

**Zur Leistungsfähigkeit von Modellversuchen im kaufmännischen
Schulwesen der Bundesrepublik Deutschland – eine
metaevaluative Studie als Beitrag zur Modellversuchsarbeit**

Inaugural-Dissertation
zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaften
der Universität Mannheim

vorgelegt von

Michael Grotheer
aus Osterholz-Scharmbeck

Osnabrück 2003

Dekan:

Prof. Dr. Martin Weber

Referent:

Prof. Dr. Hermann G. Ebner

Korreferentin:

Prof. Dr. Sabine Matthäus

Tag der mündlichen Prüfung:

06. Mai 2004

Inhalt

Abkürzungsverzeichnis	V
Verzeichnis der Abbildungen und Tabellen	VI
1 Ausgangslage	1
1.1 Problemformulierung	1
1.2 Zum Aufbau der Arbeit	6
2 Zum Stand der Auseinandersetzung mit der Modellversuchsarbeit ...	11
2.1 Zur Definition des Begriffes 'Modellversuch'	11
2.1.1 Semantische Begriffserläuterung	11
2.1.2 Organisationsbezogene (administrative) Begriffserläute- rung	13
2.1.3 Funktionale Begriffserläuterung	13
2.2 Inhaltsbereiche der wissenschaftlichen Diskussion	16
2.2.1 Transferproblematik von Modellversuchsergebnissen	17
2.2.2 Modellversuche und Organisationsentwicklung	20
2.2.3 Leistungen von Modellversuchen – Modellversuche als Träger von Innovationen für die Berufsbildungspraxis	23
2.3 Erste Zwischenbilanz: Ausprägungen der inhaltlichen Schwerpunk- te in der Modellversuchsliteratur	26
2.3.1 Auseinandersetzung mit der Transferproblematik von Mo- dellversuchsergebnissen	26
2.3.2 Auseinandersetzung mit dem Aspekt 'Modellversuche und Organisationsentwicklung'	29
2.3.3 Auseinandersetzung mit dem Aspekt der 'Leistungen' von Modellversuchen	29
3 Bildungspolitischer Rückblick und Untersuchung der finanziellen Förderungsentwicklung (Teil A der Studie)	31
3.1 Bildungspolitischer Rückblick	31
3.1.1 Ausgangslage	31
3.1.2 Systematisierung der geförderten Modellversuche	31
3.2 Ziel-Inhaltsanalyse von Berufsbildungsberichten in ihrer Funktion bildungspolitischer ex-post Dokumentationen	34
3.3 Auswertung der Untersuchung berufsbildungspolitischer Intentio- nen - 1983-1998 Zusammenfassung und Einordnung der Modell- versuchsarbeit in den berufsbildungspolitischen Horizont	35
3.3.1 Entwicklungslinien in der Berufsbildungspolitik	35
3.3.2 Kohärenz der Modellversuchsarbeit mit der Berufsbil- dungspolitik des Untersuchungszeitraumes	38
3.4 Chronologische Untersuchung der finanziellen Förderungsentwick- lung	41
3.4.1 Darstellung der finanziellen Förderungsentwicklung von BLK-Modellversuchen	42
3.4.2 Darstellungs- und Interpretationsprobleme im Hinblick auf den Input-, Outputzusammenhang	46

4	Zum Stand der Evaluationsforschung	48
4.1	Der Begriff 'Evaluation'	48
4.2	Entwicklungslinien in der Evaluationsmethodologie der nordamerikanischen Forschungstradition	54
4.2.1	Chronologische Entwicklung der Evaluationsmethodologie	54
4.2.2	Typologische Darstellung der Evaluationsansätze	55
4.3	Theoriegeleitete Evaluation	63
4.3.1	Aufbau und Dimensionen des programmtheoretischen Ansatzes von Chen	63
4.3.2	Wertorientierung im programmtheoretischen Ansatz von Chen	68
4.4	Klassifikation und Analyse von theoriebasierenden Programmevaluationsansätzen nach Stufflebeam	72
4.4.1	Darstellung des Evaluationsvorhabens	72
4.4.2	Stufflebeams Einschätzung theoriebasierender Evaluationsansätze	74
4.5	Zweite Zwischenbilanz: Der Stand der Forschung in der Evaluationsliteratur	75
5	Entwicklung des Evaluationskonzeptes	78
5.1	Konzeptspezifikation und Operationalisierung	78
5.2	Bestimmung der anzuwendenden Evaluationsstandards per Ausschlussverfahren	80
5.3	Ausgewählte Evaluationsstandards zur Anwendung im Rahmen der metaevaluativen Dokumentenanalyse	83
5.4	Kongruenzprüfung von evaluationsbezogenen Zielen und hierauf bezogene Evaluationsverfahren	88
6	Evaluation ausgewählter Modellversuche (Teil B der Studie)	91
6.1	Ausgewählte Modellversuche der Gruppe G4 – Innovationen in bestehenden Fächern oder Schulformen	91
6.2	Darstellung und Evaluation des Modellversuchs 'KUNO'	92
6.2.1	Beschreibung des Modellversuchs	92
6.2.1.1	Allgemeine und strukturelle Angaben	92
6.2.1.2	Modellversuchsziele und –inhalte/ Tätigkeiten der WBL auf Modellversuchsebene	93
6.2.1.3	Fortbildungen/ Tagungen/ modellversuchsbegleitende Transferaktivitäten	95
6.2.1.4	Ergebnisse des Modellversuchs aus Sicht der WBL	95
6.2.2	Darstellung des summativen Evaluationsansatzes der WBL im Modellversuch 'KUNO'	95
6.2.3	Ergebnis der Untersuchung – Ausprägungsgrade der Evaluationsstandards	96
6.3	Darstellung und Evaluation des Modellversuchs 'UMWELT'	102
6.3.1	Beschreibung des Modellversuchs	102

6.3.1.1	Allgemeine und strukturelle Angaben	102
6.3.1.2	Modellversuchsziele und –inhalte/ Tätigkeiten der WBL auf Modellversuchsebene	103
6.3.1.3	Fortbildungen/ Tagungen/ modellversuchsbegleitende Transferaktivitäten	105
6.3.1.4	Ergebnisse des Modellversuchs aus Sicht der WBL	105
6.3.2	Darstellung des summativen Evaluationsansatzes der WBL im Modellversuch 'UMWELT'	105
6.3.3	Ergebnis der Untersuchung – Ausprägungsgrade der Evaluationsstandards	106
6.4	Darstellung und Evaluation des Modellversuchs 'SCHLÜSSELQUALIFIKATIONEN'	110
6.4.1	Beschreibung des Modellversuchs	110
6.4.1.1	Allgemeine und strukturelle Angaben	110
6.4.1.2	Modellversuchsziele und –inhalte/ Tätigkeiten der WBL auf Modellversuchsebene	110
6.4.1.3	Fortbildungen/ Tagungen/ modellversuchsbegleitende Transferaktivitäten	112
6.4.1.4	Ergebnisse des Modellversuchs aus Sicht der WBL	112
6.4.2	Darstellung des summativen Evaluationsansatzes der WBL im Modellversuch 'SCHLÜSSELQUALIFIKATIONEN'	113
6.4.3	Ergebnis der Untersuchung – Ausprägungsgrade der Evaluationsstandards	115
6.5	Darstellung und Evaluation des Modellversuchs 'SOL'	123
6.5.1	Beschreibung des Modellversuchs	123
6.5.1.1	Allgemeine und strukturelle Angaben	123
6.5.1.2	Modellversuchsziele und –inhalte/ Tätigkeiten der WBL auf Modellversuchsebene	124
6.5.1.3	Fortbildungen/ Tagungen/ modellversuchsbegleitende Transferaktivitäten	125
6.5.1.4	Ergebnisse des Modellversuchs aus Sicht der WBL	125
6.5.2	Darstellung des summativen Evaluationsansatzes der WBL im Modellversuch 'SOL'	126
6.5.3	Ergebnis der Untersuchung – Ausprägungsgrade der Evaluationsstandards	127
6.6	Darstellung und Evaluation des Modellversuchs 'EUWAS'	135
6.6.1	Beschreibung des Modellversuchs	135
6.6.1.1	Allgemeine und strukturelle Angaben	135
6.6.1.2	Modellversuchsziele und –inhalte/ Tätigkeiten der WBL auf Modellversuchsebene	136
6.6.1.3	Ergebnisse des Modellversuchs aus Sicht der WBL	138

6.6.2	Darstellung des summativen Evaluationsansatzes der WBL im Modellversuch 'EUWAS'	138
6.6.3	Ergebnis der Untersuchung – Ausprägungsgrade der Evaluationsstandards	139
6.7	Darstellung und Evaluation des Modellversuchs 'GANZHEITLICHE LERNANSÄTZE'	148
6.7.1	Beschreibung des Modellversuchs	148
6.7.1.1	Allgemeine und strukturelle Angaben	148
6.7.1.2	Modellversuchsziele und –inhalte/ Tätigkeiten der WBL auf Modellversuchsebene	148
6.7.1.3	Fortbildungen/ Tagungen/ modellversuchsbegleitende Transferaktivitäten	149
6.7.2	Ergebnisse des Modellversuchs aus Sicht der WBL und Zwischenergebnis/Resümee	149
6.8	Datenauswertung und Interpretation	150
6.8.1	Zur Aussagefähigkeit und Darstellung der Mittelwerte der Standardbereiche	150
6.8.2	Interpretation der Mittelwerte der Standardbereiche	151
7	Zusammenfassung und Ausblick	155
7.1	Auswertung der deskriptiv-analytischen sowie der evaluativen Aspekte der Studie	155
7.1.1	Auswertung und Fazit des deskriptiv-analytischen Teils der Studie (Teil A)	156
7.1.2	Auswertung und Fazit des evaluativen Teils der Studie (Teil B)	159
7.2	Diskussionsansätze zur Steigerung der Leistungsfähigkeit des Instruments 'Modellversuch'	162
	Literaturverzeichnis	166
	Anhang zu Teil A der Studie	179
	Anhang zu Teil B der Studie	188

Abkürzungsverzeichnis

AB	Abschlussbericht des jeweiligen Modellversuchs
Abb.	Abbildung
Anm. Verf.	Anmerkung Verfasser
B&B	Beteiligte und Betroffene
BerBiFG	Berufsbildungsförderungsgesetz
BiBB	Bundesinstitut für Berufsbildung
BIP	Bruttoinlandsprodukt
BLK	Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung
BMBF	Bundesministerium für Bildung und Forschung ehemals BMBW – Bundesministerium für Bildung, Wissenschaft und Technologie
D (z. B. D1)	Durchführbarkeitsstandards
DFG	Deutsche Forschungsgemeinschaft
EUWAS	Euro-Wirtschaftsassistent
EV	Evaluation
FN	Fußnote
G (z. B. G1)	Genauigkeitsstandards
GOE	Gesellschaft für Organisationsentwicklung
HeLP	Hessisches Landesinstitut für Pädagogik
IKT	Informations- und Kommunikationstechnologien
K (z. B. K1)	Korrektheitsstandards
k. T.	kausative Theorie
Kap.	Kapitel
KUNO	Kundenorientierung in der kaufmännischen Berufsbildung
LIS	Landesinstitut für Schule – Bremen
MV	Modellversuch
N (z. B. N1)	Nützlichkeitsstandards
n. T.	normative Theorie
OE	Organisationsentwicklung
OECD	Organisation for Economic Co-operation and Development
SOGU/S	Selbstorganisationsgrad von Unterricht/Schülerfragebogen
SOL	Selbstorganisierte Lernprozesse
SQ	Schlüsselqualifikation
SUKESOL	Subjektive Kompetenzeinschätzung selbstorganisierten Lernens
Tab.	Tabelle
TeS	Transfer im engeren Sinne
TwS	Transfer im weiteren Sinne
VGR	Volkswirtschaftliche Gesamtrechnung
WBL	Wissenschaftliche Begleitung
ZB	Zwischenbericht des jeweiligen Modellversuchs

Verzeichnis der Abbildungen und Tabellen

Abbildung 1	Untersuchungsdesign zur Ermittlung der Leistungsfähigkeit von Modellversuchen	10
Abbildung 2	Begriffsfassung Modellversuch	11
Abbildung 3	Gruppen von der BLK geförderter Modellversuche	34
Tabelle 1	Bildungspolitische Schwerpunkte.....	37
Tabelle 2	Initiierungsjahre ausgewählter Modellversuche	39
Abbildung 4	Entwicklung der Bildungsausgaben für BLK-Modellversuche	42
Tabelle 3	Entwicklung der Fördermittel des Bundes für BLK-Modellversuche	43
Tabelle 4	Ausgewählte Modellversuche der Gruppe 4: Innovationen in bestehenden Fächern oder Schulformen	(Anhang zu Teil B) 202
Abbildung 5	Ausgaben für Wissenschaft und Forschung außerhalb der Hochschulen	44
Abbildung 6	Anteil der Bundes-Modellversuchsausgaben an den Ausgaben für Wissenschaft und Forschung außerhalb der Hochschulen	45
Abbildung 7	Evaluand-Wirkungs-Matrix	51
Abbildung 8	Abgrenzung von Evaluation und Bildungsforschung	54
Abbildung 9	Phasen der Evaluationsmethodologie nach Beywl	55
Abbildung 10	Programmtheoretischer Ansatz nach Chen	64
Abbildung 11	Gegenüberstellung der 'Werte' des quantitativen und des qualitativen Forschungsparadigmata	69
Abbildung 12	Kongruenzprüfung von Evaluationszielen und Evaluationsverfahren	90
Tabelle 5	Ergebnis der metaevaluativen Untersuchung MV KU-NO/Bremen	96
Tabelle 6	Ergebnis der metaevaluativen Untersuchung MV Umwelt/Bremen	106
Tabelle 7	Ergebnis der metaevaluativen Untersuchung MV Schlüsselqualifikationen/Hamburg	115
Tabelle 8	Ergebnis der metaevaluativen Untersuchung MV SOL/Hessen	127
Tabelle 9	Ergebnis der metaevaluativen Untersuchung MV EUWAS/Nordrhein-Westfalen	139
Tabelle 10	Mittelwerte der Evaluationsstandards	151

1 Ausgangslage

1.1 Problemformulierung

Das ‚Neue Globale Zeitalter‘, so die OECD, die damit ihren Fokus auf die Zukunft im Jahr 2020 richtet, ist gekennzeichnet durch eine höhere Prosperität bei gleichzeitigem Wettbewerbsdruck und rascheren Strukturwandel.

Veränderungen in wirtschafts- und bildungspolitischen Bereichen haben in den vergangenen zehn Jahren unter dem Einfluss marktorientierter Wirtschaftsentwicklung, dem raschen Wirtschaftswachstum verschiedener nicht-OECD-Volkswirtschaften, dem technischen Fortschritt, vor allem auf dem Gebiet der Informations- und Kommunikationstechnologien, und anderer Faktoren zugenommen (vgl. OECD 1998: 15). Diese Entwicklung erfordert Flexibilität, vor allem auch auf dem Arbeitsmarkt, für den der Bildungsbereich die (zukünftigen) Arbeitnehmer zu qualifizieren hat. In diesem Zusammenhang besteht die Gefahr, dass ein Stillstand einzelner Faktoren (allgemeine Bildung, berufliche Bildung, Hochschulbildung) ein Rückstand der genannten Sektoren (Arbeitsmarkt und Bildungsbereich) im internationalen Kontext insgesamt bedeuten kann. Für den beruflichen Bildungsbereich bedeutet dies, dass Innovationen unerlässlich sind, um Arbeitnehmerinnen und Arbeitnehmer auf künftige Entwicklungen vorzubereiten (vgl. Heimerer 1998: 326f).

Doch wodurch werden Entwicklungen im Bereich der Bildung, speziell der beruflichen Bildung initiiert? Als Nachfrager von Arbeitskraft sind die Unternehmen und natürlich auch die Anbieter der Arbeitskraft, die Arbeitnehmer, daran interessiert, sich konsequent und adaptiv auf die Zukunft, auf die 'Wissens- und Informationsgesellschaft' vorzubereiten.

Dehnbostel formuliert die nach seiner Meinung notwendigen Anforderungen hierzu wie folgt: "Ein hohes Qualifikationsniveau ist die Voraussetzung für die wirtschaftlich-technische und ökologische Entwicklung, für internationale Konkurrenzfähigkeit und für eine zukunftsgerichtete Flexibilität des Arbeitsmarktes. Moderne Berufsausbildung und ständige berufliche Weiterbildung stärken die berufliche Kompetenz und Autonomie des Einzelnen sowie die Innovationsfähigkeit der Wirtschaft" (Dehnbostel 1994: 4).

Unterstützung erfährt die Aussage durch Ergebnisse von Untersuchungen aus dem Bereich der Bildungsökonomie. Diese interdisziplinäre Forschungsrichtung, hat die Korrelation von Bildungsstand einer Volkswirtschaft und deren wirtschaftlichen Erfolg nachgewiesen. Widmaier und Frey entwarfen hierzu ein globales Modell der Bildungsplanung, das "[...] Interdependenzen zwischen Bildungs- und Ausbildungssystem und wirtschaftlicher Entwicklung (aufzeigt; Anm. Verf.) und greifbare Hinweise für praktisch notwendig werdende Entscheidungen und die Möglichkeit ihrer Verwirklichung [...]" (Widmaier; Frey 1972: 93ff) gibt. Das Ergebnis sind Trendprognosen, die einen Beitrag zur politischen Entscheidungsfindung leisten. Inwieweit ein Modell, als 'reduziertes Abbild' der Realität, wie es in der Bildungsökonomie entwickelt wird, mehr als die von ihm erwartete Erklärungsfunktion wahrnehmen kann, sondern darüber hinaus praktisch umsetzbare Hinweise für offene

Handlungsfelder entwickelt, erscheint problematisch. Unbestritten ist jedoch, dass das im Bildungssystem ausgebildete Humankapital im Wirtschaftsprozess neben dem Sachkapital eine bedeutende Größe darstellt.

Der in der Bildungsökonomie enthaltene Verweis auf produktive und konsumptive Bestandteile von Bildung hebt außerdem diejenigen Bereiche hervor, für die der Bildungsprozess Leistungen erbringen kann. In welchem Zusammenhang die Ergebnisse der Bildung produktiven oder konsumptiven Charakter einnehmen, hängt vom spezifischen Einsatzgebiet der erworbenen Qualifikationen ab. In einem Wirtschaftssystem jedoch, das sich ständig den wandelnden Anforderungen der Konsumenten anpasst, kann Wissen und Denken, das ursprünglich nicht der 'Utilität', im Sinne der beruflichen Verwertbarkeit diene, schnell einen hohen Stellenwert in der technisch-ökonomischen Welt erlangen und ein in diesem Zusammenhang verwertbares Produkt darstellen (vgl. Recum v. 1972: 49ff). Durch den 'Utilitäts-Gedanken' wird Bildung auch als Investition für den Einzelnen bzw. des Einzelnen betrachtet, die Kosten verursacht, gleichzeitig aber auch zukünftige Erträge bewirkt, da sie im Erwerbsprozess eingesetzt werden kann (vgl. Helberger, Palamidis, 1989: 209ff).

Vorausgesetzt, die Kultusminister der Länder wollen die zur Anpassung an die ökonomisch-technische Entwicklung notwendigen Änderungen oder über das Maß der Anpassung hinaus bedeutend gewordene Neuerungen nicht lediglich vorschreiben bzw. Bildungsgänge oder didaktisch-methodische Ansätze nicht unerprobt integrieren und damit die Gefahr eines Misserfolges riskieren, bedarf es hierzu eines entsprechenden Instrumentariums/eines Problemlösungsansatzes: Eine Möglichkeit, Innovationen anzuregen und zu erproben, stellen die in der schulischen und außerschulischen Bildung durchgeführten Modellversuche dar:

Modellversuche "[...] dienen der Entwicklung und Erprobung neuer inhaltlicher, struktureller und organisatorischer Ansätze im Bildungswesen. [...] Die wissenschaftliche Begleitung hat die Aufgabe, die Planung und Durchführung der Modellversuche zu unterstützen, ihre Ergebnisse zu beschreiben und ihre Auswirkungen zu analysieren" (BLK 1981: 7).

Nach Angaben von Knauss hat es Erprobungen und Modelle im Bildungswesen auch schon vor dem zweiten Weltkrieg gegeben (vgl. Knauss 1978: 221). Doch mit Ablauf der 60er-Jahre ergab sich eine wesentliche Änderung: Neben den praktischen Erfahrungen sollten vor allem auch wissenschaftlich gesicherte Erkenntnisse gewonnen werden.

Die Obliegenheiten, die man mit diesem Instrument verbindet, beziehen sich seit der BLK-Gründung 1970 auf drei Aufgaben: *Erstens*, so die ursprüngliche Intention, dienen sie der Politikberatung und leisten (*zweitens*) einen Beitrag zur Bildungsplanung; Modellversuche "[...] wurden als eines der wichtigsten Instrumente innovativer Bildungspolitik und Bildungsplanung konzipiert" (Dehnbostel 1998: 188). Durch sie können geplante Veränderungen im Bildungsbereich, wie neue Unterrichtskonzepte oder Schulformen, theoriegeleitet erprobt und die Ergebnisse für die weitere Planung genutzt werden. Der *dritte* Aufgabenbereich bezieht sich auf die

wissenschaftliche Begleitung, die sich durch ihre Dokumentationen und Analysen als Teil der (Berufs-) Bildungsforschung manifestiert (vgl. Sommer 1996: 22).

Mögliche Überforderungen dieser Erprobungsvorhaben ergeben sich aus der genannten Trilogie der Aufgaben und dem hohen Anspruchsniveau, das mit ihnen verbunden wird: "Durch Modellversuche soll einerseits Bestehendes fortentwickelt, andererseits Neues erprobt werden. Sie sollen so ausgerichtet sein, dass sie wichtige Entscheidungen für die Entwicklung des Bildungswesens geben" (BLK 1976: 4).

Mein Vorhaben stützt sich auf die Arbeitsthese, dass Modellversuche die Ansprüche im Hinblick auf Innovationskraft und Veränderungspotential für das Praxisfeld, wie sie schon zu Beginn der 70er-Jahre formuliert wurden, nicht erfüllen können. Dabei soll nicht bestritten werden, dass Modellversuche gewisse Leistungen erbringen, aber ob sie 'wichtige Entscheidungshilfen' für die Entwicklung des Bildungswesens geben, sei dahingestellt.

Dieser These soll nachgegangen werden, indem ausgewählte Modellversuche einerseits unter deskriptiv-analytischen Aspekten betrachtet und andererseits einer metaevaluativen Untersuchung unterzogen werden. Der erste Teil der Arbeit beschäftigt sich daher auch zunächst mit vorliegenden 'harten' Daten (deskriptiver Teil - A): der finanziellen Förderungsentwicklung sowie einem bildungspolitischen Rückblick. Darüber hinaus beinhaltet der zweite Teil (Teil B) die Entwicklung und (metaevaluative) Anwendung eines Evaluationsrasters auf Basis anerkannter und hinreichend elaborierter Evaluationsstandards.

Bisher ist die Auseinandersetzung mit den Leistungen des Instruments Modellversuch gering, d. h. bisher wurden keine Anstrengungen unternommen (siehe auch Kap. 2), operationale Messkriterien zu entwickeln oder einen einheitlichen 'Leistungsstandard' zu definieren. Es wurde unterlassen, Modellversuchsergebnisse fassbar zu machen, was die Gewinnung von Erkenntnissen über die 'Leistungsfähigkeit' der Erprobungsvorhaben blockiert (vgl. Kap. 2.3.3).

Die folgenden Ausführungen zu den Aspekten Berufsbildungspolitik und Berufsbildungsplanung, für die Modellversuche Leistungen erbringen, stellen Begründungsdimensionen für die Auseinandersetzung mit dem gewählten Themenbereich der Modellversuchsarbeit bzw. Modellversuchspolitik dar und verdeutlichen, vor allem in Zusammenhang der Auseinandersetzung mit der Modellversuchsarbeit in Kapitel 2, die Dringlichkeit der Beschäftigung mit der Thematik:

Eine gemeinsame Bildungsplanung von Bund und Ländern gibt es, trotz föderalistischem System und der damit verbundenen Kulturhoheit der Länder, seit 1970 mit Aufnahme der Tätigkeit der Bund-Länder-Kommission für Bildungsplanung (seit 1975 ist diese auch für die Forschungsförderung zuständig), nachdem zuvor im Herbst 1969 das Bundesministerium für Bildung und Wissenschaft (BMBW) gegründet wurde. Mit Konstitution der BLK wurde ein Ausschuss für Innovationen im

Bildungswesen eingesetzt, der Vorschläge für gemeinsame Innovationsprojekte der Bildungsplanung und Bildungsforschung von überregionaler Bedeutung erarbeitete. Ziel war es, zunächst eine Vereinbarung über Vorbereitung, Durchführung und wissenschaftliche Begleitung von Modellversuchen im Bildungswesen zu entwerfen. Diese wurde 1971 von den Regierungschefs von Bund und Ländern unterzeichnet und stellt die Rechtsgrundlage für das System der Modellversuchsförderung dar (vgl. Knauss 1978: 221ff).

Durch diese Vereinbarung dokumentieren Bund und Länder ihren politischen Willen zur gemeinsamen Umsetzung von Planvorstellungen. Hierin ist die Aussage enthalten, dass diese Art von Erprobungsmaßnahmen als Instrument zukünftigen Wandels wichtig und notwendig ist und als Element eines Reformprozesses angesehen werden kann. Daraus folgt auch, dass Entscheidungen auf einem höheren Erkenntnisniveau getroffen und die Gefahr von Fehlentscheidungen verringert werden soll. Unberücksichtigt bleibt dabei der monetäre Faktor, der dieser Vereinbarung noch eine andere Sicht verleiht. Durch die Bezuschussung behält sich das jetzige Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF; vormals BMBW) die Mitentscheidung über bildungspolitische Entwicklungen innerhalb der Länder vor.

Es liegt auf der Hand, dass das BMBF den so gewonnenen Einfluss nicht missen will, damit einheitliche Bildungspolitik möglich wird. Aus Sicht der Länder könnte die Modellversuchspolitik als Refinanzierungsmöglichkeit gesehen werden, was sich aus den Förderungsmodalitäten der Modellversuche ergibt. Die BLK übernimmt 50 % der erforderlichen besonderen Mittel, die andere Hälfte wird vom antragstellenden Bundesland aufgebracht (vgl. BLK 1997: 9). Dieses hat wiederum die Möglichkeit, lt. vorgeschriebenem Finanzierungsplan der BLK, auch Personalkosten mit einzubeziehen. Unter dem Gesichtspunkt der Internalisierung von Kosten ist dieses Vorgehen sicherlich richtig. Doch eine rein betriebswirtschaftliche Kostenrechnung verschleiern, dass durch die Berücksichtigung der Personalkosten wichtige Finanzmittel nicht real aufgebracht und somit in absoluter Höhe nicht für das geplante Vorhaben zur Verfügung gestellt werden müssen. Es besteht außerdem die Gefahr, dass die Bundesmittel für Investivkapital (Schulausstattung) ausgegeben werden, was eine hohe Kapitalbindung mit sich bringt; Folgekosten hier nicht mitbetrachtet.

Es wird deutlich, dass schon die Förderungsmodalitäten eines bildungspolitischen Instruments, wie es Modellversuche darstellen, Elemente enthalten, die zu einer Kritik an der Wirksamkeit von Modellversuchen insgesamt führen können.

Generell, d. h. über die o. g. Aufgaben-Trilogie von Modellversuchen sowie politisch-strategischer Überlegungen hinaus, gilt es jedoch festzustellen, dass Modellversuche ihre Bedeutung (und die Auseinandersetzung mit ihnen) Brisanz durch einen Faktor erhalten, der Bildungsplanung und somit auch Bildungspolitik ohne sie unsicher werden lässt. Gemeint ist die relative Unvorhersehbarkeit zukünftiger Entwicklungen. Vor diesem Hintergrund wird davon ausgegangen, dass Modellver-

suche diesen Faktor minimieren können, indem neue Fächer oder Innovationen in bestehenden Fächern usw. zunächst erprobt werden. Hauptbestreben ist es, sich auf diesem Wege den globalen, d. h. internationalen und nationalen Entwicklungen anpassen zu können (vgl. Ploghaus 1995: 104). Ob das Instrument 'Modellversuch' diese Leistungen erbringen kann, ist bislang noch nicht untersucht worden.

Bezieht man die sich ständig schneller verändernden Anforderungen an das Bildungswesen mit in die Betrachtung ein, dann ist es trotz möglicher Schwierigkeiten (Aufgabenvielfalt u. a., s. o.) offensichtlich, dass die Zusammenarbeit von Politik und Wissenschaft notwendig ist, um gemeinsam Ziele, Strukturen, Umfang und Qualität von Bildungseinrichtungen derart in die Zukunft zu projizieren, dass die Schulen den Bildungsansprüchen des Einzelnen und dem Gebot der Chancengleichheit einerseits und den Leistungsanforderungen der Gesellschaft (und damit auch der Wirtschaft als Subsystem der Gesellschaft) andererseits gerecht werden können.

Aus den persönlichen und gesellschaftlichen Ansprüchen an die Schulen resultiert der Bedarf an bestimmten Qualifikationen, die Nachfrage nach Bildungsabschlüssen aus der Bevölkerung sowie die Frage nach den Kosten zur Realisierung dieser Entwicklung. Faktoren, die Einfluss auf künftige Schwerpunkte in der *Bildungsplanung* nehmen und für die Modellversuche einen Entwicklungsbeitrag leisten können (vgl. Schmiel, Sommer 1985: 22).

Staatliche *Berufsbildungspolitik* und *Berufsbildungsplanung* zielen darauf ab, mit öffentlichen Mitteln die Leistungskraft und Qualität des Berufsbildungssystems zu steigern; diesem Ziel ist auch die Modellversuchförderung verpflichtet. Betrachtet man die Finanzsumme, die seit 1971 für diese Art pädagogischen Experimentes zur Verbesserung der Berufsbildung ausgegeben wurde (es handelt sich um etwa 0,9 Mrd. € für die Bereiche Schule, Berufliche Bildung und Hochschule, die allein vom Bund aufgebracht wurden (vgl. BLK 1995: 3)), dann wird deutlich, dass sich bildungspolitische Entscheidungen auf die Zuverlässigkeit und Übertragbarkeit der Erprobungen verlassen können müssen, wenn diese Ressourcen nicht wirkungslos bleiben sollen.

Der Stellenwert von Modellversuchen für die Bildungspolitik und –planung scheint jedoch nicht vollständig geklärt. Sind es Elemente einer transparenten Bildungspolitik, die sich der Beteiligung der betroffenen Gesellschaftsgruppen (Schulpraktiker u. a.) verpflichtet fühlt und die zu einer Verbesserung des Bildungswesens beitragen (politisch-handelnde Ebene) oder handelt es sich vielmehr um Instrumente, die durch den längerfristigen Erprobungszeitraum zu einer Verschleppung notwendiger Reformen führen können (politisch-strategische Ebene)? Es lässt sich die These aufstellen, dass die Funktionen von Modellversuchen als Elemente von Bildungspolitik und Bildungsplanung nicht präzise definiert sind. Eine Untersuchung der Kohärenz von Modellversuchsinisierung und berufsbildungspolitischer Zielformulierung, soll diese Arbeitsthese klären (siehe Kap. 3.3.2).

1.2 Zum Aufbau der Arbeit

Gegenstand dieser Arbeit ist die Untersuchung der 'Leistungsfähigkeit' von Modellversuchen für die Berufsbildungspraxis.

Hierzu erfolgt (in Teil B¹) eine Konzentration auf den Zeitrahmen 1971 bis 1998, da die hierin eingebundenen Modellversuche in Bezug auf die zu dieser Zeit geltenden organisatorischen Rahmenbedingungen (Einteilung von Modellversuchen in 'Förderbereiche') eine abgrenzbare und homogene Gruppe darstellt. Ab 1998 hat die BLK damit begonnen, eine 'Programmförderung' innerhalb festgelegter Förderschwerpunkte zu initiieren. Die Programme fassen jeweils Fragestellungen innerhalb eines Schwerpunktes zusammen, die länderübergreifend für die Weiterentwicklung des Bildungswesens von Bedeutung sein sollen. Ziel ist es, durch Programme breiter abgesicherte Entscheidungshilfen zu gewähren.

Die Aussagefähigkeit der durch diese Arbeit zu generierenden Erkenntnisse nimmt trotz veränderter Rahmenbedingungen jedoch nicht ab, da die Qualität der Leistungen von Modellversuchen insbesondere durch die Evaluation der wissenschaftlichen Begleitung sicher gestellt wird, an die durch die Neuordnung der Förderbedingungen keine veränderten Anforderungen gestellt werden. Die BLK hebt lediglich hervor, dass Modellversuche "[...] systematisch zu analysieren seien" (vgl. BLK: <http://www.blk-bonn.de/modellversuche/mv-neuordnung.htm>, 21.02.2003). Die gewonnenen Erkenntnisse sind somit auch auf den Zeitraum nach 1998 zu übertragen, resp. erhalten eine größere Bedeutung, da durch die Programmförderung eine Konzentration von Interessenschwerpunkten erfolgt, so dass die Aussagefähigkeit von Programm-Evaluationen eine größere Tragweite erhalten als die Evaluation von Einzelvorhaben.

Neben der Eingrenzung der chronologischen Dimension erfolgt eine inhaltliche Fokussierung der Arbeit auf zwei Untersuchungsbereiche:

Teil A beinhaltet zum einen eine Analyse von Berufsbildungsberichten, die die bildungspolitischen Intentionen der im Bezugszeitraum amtierenden Bundesbildungsminister in Bezug auf das Instrument 'Modellversuch' aufzeigen soll, um so Aussagen über den Stellenwert der Erprobungsvorhaben tätigen zu können und um zu prüfen, inwieweit Modellversuche initiiert wurden, die die jeweilige berufsbildungspolitische Intention unterstützt haben. Hierzu wird ein Vergleich von zu operationalisierenden Berufsbildungszielen und deren Erreichung anhand der berufsbildungspolitischen ex-post Dokumentationen, den Berufsbildungsberichten, vorgenommen.

Darüber hinaus soll die Analyse der finanziellen Größenordnungen der Förderung weiteren Aufschluss über die bildungspolitische Bedeutung von Modellversuchen geben, da finanzielle Größenordnungen der Förderung als Indiz für die Bedeutung betrachtet werden können, die einem Instrument wie den Modellversuchen zugemessen wird. Wenn die absoluten Beträge der finanziellen Förderung abnehmen, sind zunächst zwei Folgerungen denkbar: zum einen besteht in dem jeweili-

1) In Teil A muss die Untersuchung auf den Zeitrahmen 1982 (1983) bis 1998 beschränkt bleiben, da lt. Auskunft der BLK das Basismaterial für die Zeit vor 1982 nicht mehr vorliegt.

gen Bereich kein praktischer oder wissenschaftlicher Handlungsbedarf mehr oder aber es erfolgte eine Prioritätenverlagerung der politischen Entscheidungsträger zu Gunsten anderer Bereiche. Denkbar ist sicherlich auch, dass insgesamt Finanzknappheit der öffentlichen Kassen besteht, so dass Ausgaben reduziert werden müssen oder sich die Voraussetzungen der ‚Geldbeschaffung‘ durch Personalwechsel in der BLK oder dem BiBB verändern.

Um die Bedeutsamkeit von Modellversuchen zu bestimmen, wird die Entwicklung der finanziellen Größenordnungen der Förderung anhand der Bildungsausgaben für diesen Bereich nachvollzogen. Als Vergleichsmaßstab zu diesen Größen werden den absoluten DM-Beträgen die Entwicklung des Brutto-Inlands-Produkts (BIP) sowie andere Bezugsgrößen entgegengestellt, um deren relative Höhe zu ermitteln.²

Teil A stellt demnach eine Auseinandersetzung mit den 'leistungsbeeinflussenden' Rahmenbedingungen von Modellversuchen auf einer bildungspolitisch-finanziellen Ebene dar (deskriptiver-analytischer Teil), während in **Teil B** die Auseinandersetzung mit den Leistungen ausgewählter Modellversuche auf (meta-) evaluativer Ebene durch Anlehnung von Evaluationsstandards an Zwischen- und Abschlussberichten von Modellversuchen erfolgt. Dieses Vorgehen kann als (metaevaluative) Sekundäranalyse von Modellversuchsdokumentationen beschrieben werden.

Hierzu ist es jedoch von Bedeutung, die im Untersuchungszeitraum durchgeführten Modellversuche zu gleichartigen Gruppen zusammenzufassen (Typisierung), da die BLK bis 1998 auf eine Schwerpunktsetzung innerhalb der bestehenden Förderbereiche verzichtet hat (s. o./FN 3) und die durchgeführten Modellversuche somit eine unstrukturierte Menge mit unterschiedlichen Inhaltsdimensionen und Intentionen darstellen.³

Die Auseinandersetzung mit der Evaluationsforschung unter dem Fokus der in der einschlägigen Literatur schwerpunktmäßig vertretenen Aspekte des Gegenstandes von Evaluation und der methodologischen Diskussion (Zum Stand der Eva-

2) Auf eine Umrechnung in die gültige Eurowährung wird auf Grund der rückblickenden Betrachtung verzichtet.

3) Vor dem 1. Januar 1998 hat man lediglich Förderbereiche festgelegt, wie den der Beruflichen Bildung, die nicht systematisiert wurden (s. o. und BLK 'Informationen über Modellversuchsförderung durch Bund und Länder im Rahmen der BLK ab 01.01.1998.' <<http://www.k.shuttle.de/blk/mv-neuordnung.htm>> (18.08.1998)).

Durch die 'bildungs- und wirtschaftspolitische Entwicklungsdynamik' hat sich die BLK jedoch veranlasst gesehen, ihre Förderrichtlinien mit Wirkung zum 01.01.1998 zu ändern und zentrale Problembereiche zu formulieren, für die ein bildungspolitischer und fachlicher Handlungsbedarf besteht. Hierbei werden Schwerpunkte festgelegt, die durch Programme umgesetzt werden und eine Laufzeit von fünf Jahren umfassen.

Zur Steigerung der Transparenz des weiteren Vorgehens erfolgt im Vorgriff auf das Ergebnis der Typisierung der Hinweis, dass, zur Erhöhung der Aussagekraft der Arbeit, eine Konzentration auf schulische Modellversuche im Bereich der kaufmännischen Berufsausbildung erfolgt, mit denen das Ziel verfolgt wird, den Erwerb von Handlungskompetenz zu ermöglichen. Dieser Modellversuchstyp stellt die am häufigsten durchgeführte Art dar (vgl. Kap. 3.1.2).

lationsforschung; siehe Kap. 4) bildet die Grundlage für die Entwicklung und Legitimation des in dieser Arbeit Anwendung findenden Untersuchungsansatzes (Metaevaluation; Kap. 5). Das Ergebnis hieraus ist ein Evaluationsraster, auf anerkannten Evaluationsstandards basierend, das an Modellversuchsevaluationen (auf der Metaebene) angelehnt wird, um so die Leistungen⁴ der jeweiligen Erprobungsvorhaben zu prüfen. Als Untersuchungsgegenstand dienen dabei Zwischen- und Abschlussberichte ausgewählter Modellversuche, die daraufhin untersucht werden, ob die Evaluationen der Erprobungsvorhaben von einer Güte sind, dass ihre Dokumentationen, in denen sich die Leistungen der Evaluationen manifestieren, als transferfähige Produkte angesehen werden können.

Grundlegend für den Aufbau des Teils B der Arbeit sind folgende Ausführungen/Prämissen, die ihren Fokus auf den Aspekt der Evaluation richten:

Es wird ein Verständnis der Funktion der wissenschaftlichen Begleitung (WBL) von Modellversuchen zu Grunde gelegt, das sich auf die von ihr wahrzunehmenden (gleichwertigen)⁵ Aufgaben bezieht. Dieses Aufgabenspektrum umfasst die Planung bzw. Konstruktion, Durchführung und Evaluation von Modellversuchen (vgl. BLK 1981: 7). Insofern kann die wissenschaftliche Begleitung und deren Ergebnis, die Evaluation, als fester Bestandteil eines Modellversuches angesehen werden, deren Qualität mit dem Innovationsgehalt für das Praxisfeld des Erprobungsvorhabens insgesamt korreliert.⁶ Ist die Evaluation im Hinblick auf ihre Güte (Zielformulierung, Verfahrensdokumentation u.v.m.) zu kritisieren, dann können keine fundierten Aussagen zur Verwendung der Ergebnisse im Praxisfeld getroffen werden und das Modellhafte am Modellversuch geht verloren.

Zur Verdeutlichung der vorstehenden Ausführungen wird die von Rossi und Freeman verwendete Auslegung von 'Evaluationsforschung' herangezogen. Diese definieren den Begriff als "[...] systematische Anwendung empirischer Forschungsmethoden zur Bewertung des Konzepts, des Untersuchungsplanes, der Implementation und der Wirksamkeit sozialer Interventionsprogramme" (Rossi; Freeman 1993).

Die Autoren induzieren demnach ein anwendungsbezogenes Verständnis des Begriffes Evaluationsforschung, das, ähnlich wie bei Caporaso/Roos 1973, Cook/Levinton u. a., auf die Überprüfung der Auswirkung politisch-administrativer Maßnahmen ausgerichtet ist (vgl. Schnell 1999: 219). Um dieser Definition gerecht zu werden ist es jedoch erforderlich, dass die mit einer Maßnahme, hier der Modellversuche, zu erreichenden Ziele genau beschrieben und geprüft werden und es ist außerdem zu klären, ob die gewünschte Evaluation das methodische Instrumen-

4) Zur Definition des Leistungsbegriffes siehe Kap. 2.2.3.

5) Bei der Analyse von MV-Dokumentationen entsteht häufig der Eindruck, die Mitarbeiter der WBL gewichten die Aufgabenbereiche unterschiedlich. Eine solche Prioritätenbildung lässt sich aus den Anforderungen der BLK jedoch nicht erkennen.

6) Da die Evaluation fester Bestandteil eines Modellversuches ist (vgl. Kap. 5.1) und diese den Aufbau des Erprobungsvorhabens inklusive der erwarteten Ziele prüft und die Implementation des entwickelten Produkts (Curriculum o. ä.) begleitet, übernimmt die Evaluation eine 'qualitäts-sichernde' Funktion.

tarium erforderlich macht, das der Evaluator einsetzt (Kongruenzprüfung; siehe Bortz 1995: 103; Schnell 1999: 121ff).

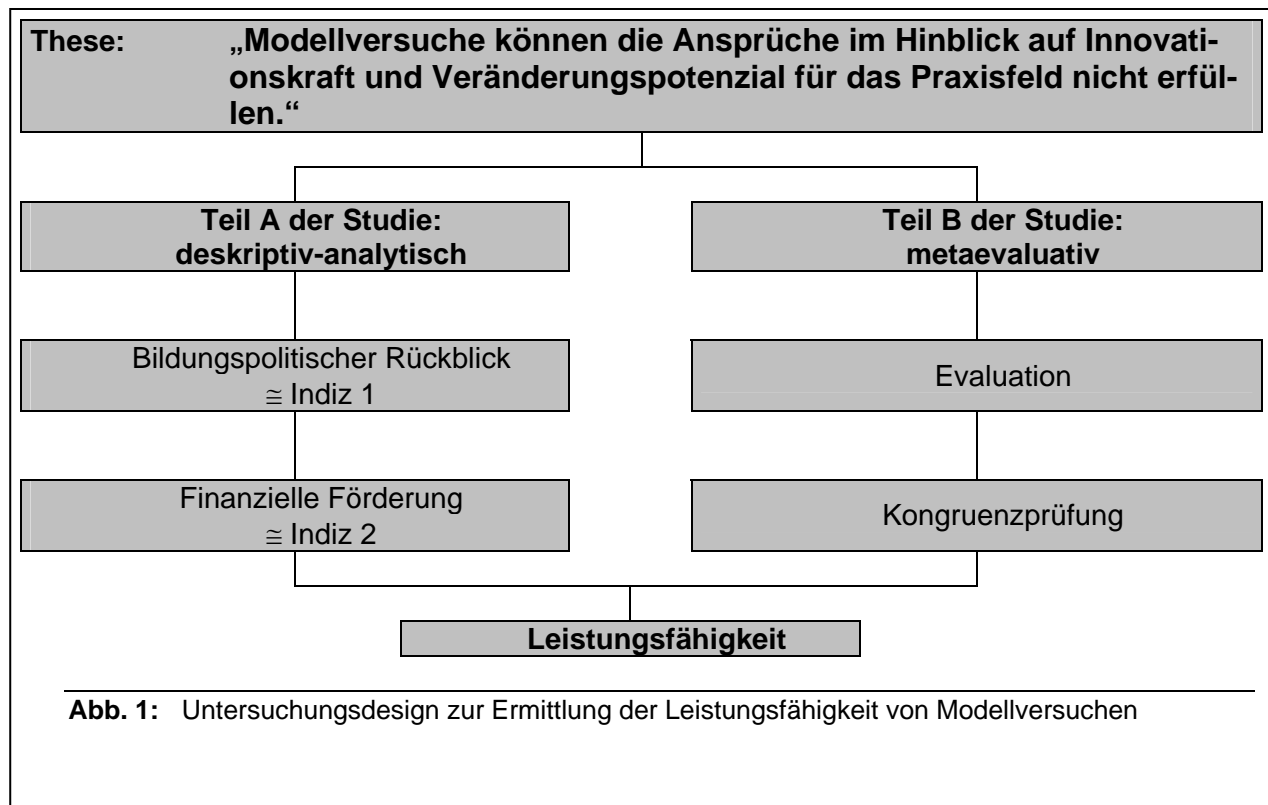
Die evaluative Aufgabe der wissenschaftlichen Begleitung bezieht sich demnach auf zwei Ebenen: zum einen auf die Evaluation des Produktes auf der Objekt-ebene, zum anderen ist auf einer Metaebene die Eignung der eingesetzten Instrumente und Methoden der Modellversuchsentwicklung und der anschließenden Evaluation auf Objektebene zu prüfen. Die Evaluation, resp. das Ausmaß, in dem sie den an sie gestellten Evaluationsstandards genügt, gibt demzufolge Aufschluss über die Leistung des einzelnen Modellversuchs. Die Erkenntnisse bezogen auf die Leistungen von Modellversuchen (s. o.) werden also auf der ‚meta-‘ evaluativen-Ebene gewonnen, die Leistungen, die internen und externen Effekte, werden jedoch (auch) auf der objektbezogenen Ebene wirksam. In Zusammenhang mit den im deskriptiven Teil der Arbeit (Teil A) ermittelten Indizien, sind hiernach Rückschlüsse zur ‚Leistungsfähigkeit‘ des ‚Instruments Modellversuch‘ insgesamt möglich.

Das Leistungsprüfverfahren gliedert sich in folgende Schritte:

1. Entwicklung eines Evaluationsrasters auf Basis elaborierter Evaluationsstandards.
2. Anwendung des Evaluationsrasters an Zwischen- und Abschlussberichte auf der Metaebene.
3. Überprüfung der Kongruenz von Zielen der Evaluation und eingesetzten Evaluationsverfahren.

Die Ergebnisse des Teils B der Arbeit beziehen sich also auf Aussagen zur Güte der Evaluation ausgewählter Modellversuche in kaufmännischen beruflichen Schulen. Auf Basis dieser Erkenntnisse sind (zumindest) Thesen über die generellen Leistungen von Modellversuchen möglich: Ist die Qualität (Güte) einer Evaluation zu kritisieren, dann ist dieser Aufgabenbereich des Modellversuches nicht oder nicht vollständig erfüllt, so dass die Mängel, die auf der Ebene der Evaluation festgestellt werden, ihre Wirkung auch auf der Objektebene zeigen (s. o.). Darüber hinaus lassen sich Empfehlungen für die Förderungspolitik formulieren, die sich einerseits in Anforderungen für die Evaluationskonzepte der wissenschaftlichen Begleitung ausdrücken, so dass mögliche ‚Schwachstellen‘ vermieden und die mit Modellversuchen verfolgte Qualität und Quantität an internen und externen Effekten gesteigert werden kann; andererseits sind Aussagen zur Notwendigkeit einer ‚angemessenen‘ inhaltlichen und finanziellen Berücksichtigung der Modellversuche als bildungspolitisches Instrument möglich.

Die nachfolgende graphische Darstellung fasst den Aufbau des Untersuchungsansatzes zusammen:



2 Zum Stand der Auseinandersetzung mit der Modellversuchsarbeit

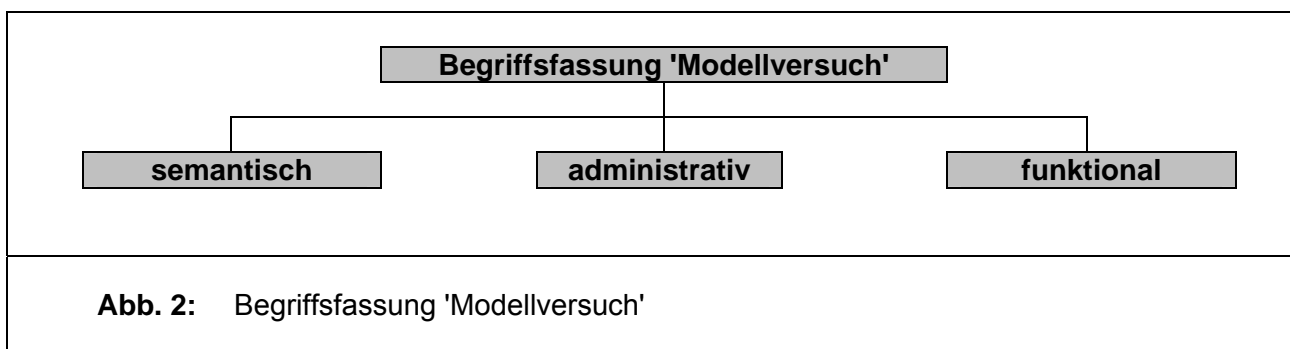
2.1 Zur Definition des Begriffes 'Modellversuch'

Da der Begriff 'Modellversuch' in der Literatur in verschiedenen Kontexten Anwendung findet, bedarf er einer Bestimmung im Hinblick auf der in dieser Arbeit verfolgten Fragestellung. Die unterschiedliche Verwendungsweise (funktional, administrativ und semantisch) der Begrifflichkeit erklärt sich aus dem jeweiligen Fokus (Schwerpunkt) des Betrachters:

So können Modellversuche unter dem Blickwinkel einer funktionalen Erklärungs- und Verwendungsvariante als Instrumente der Bildungsplanung und Bildungspraxis angesehen werden, als Erkenntnisgegenstand der Bildungsforschung oder als Element der Organisationsentwicklung.

Die zweite Herangehensweise bei der Begriffsfassung ist die Zuordnung von Modellversuchen zu ihrer finanziellen und organisatorischen Trägerschaft. Generell werden zwar alle Modellvorhaben durch das BMBF gefördert, unterliegen aber unterschiedlichen administrativen Hoheiten. Gemeint ist die Zuständigkeit der Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung einerseits (Schulmodellversuche) und des Bundesinstituts für Berufsbildung (Wirtschaftsmodellversuche) andererseits (vgl. Dehnbostel 1998: 188).

Eine dritte Möglichkeit, den Begriff 'Modellversuch' zu fassen, ist die semantische Herleitung. Dabei wird das hier vorliegende zusammengesetzte Nomen ‚Modell-Versuch‘ analysiert, d. h. es wird in seine Bestandteile zerlegt und auf seinen Bedeutungsgegenstand hin untersucht. Der Vorteil dieser Vorgehensweise liegt in der Aufdeckung sprachlicher Widersprüche, welche zu Fehlinterpretationen und damit zu falschen Erwartungshaltungen führen können.



2.1.1 Semantische Begriffserläuterung

Stachowiak bezieht sich in seinen Ausführungen zum Modellkonzept auf die systematisch-neopragmatische Erkenntnislehre und hebt hervor, dass Modelle zwar immer 'Modelle von etwas' sind, aber i. d. R. nicht alle Originalattribute erfassen, sondern nur diejenigen, die für den Verwender des Modells von Bedeutung sind, d. h. sie haben eine Ersetzungsfunktion:

- für bestimmte Erkenntnis- und Aktionssubjekte,
- innerhalb bestimmter Zeitintervalle der Originalrepräsentation,
- relativ zu bestimmten Zwecken und Zielen, denen die Modellbildung unterliegt (vgl. Stachowiak 1983: 118).

Ein Modell wird demnach durch mindestens fünf Attribute beschrieben: "X ist Modell des Originals Y für den Verwender v in der Zeitspanne t bezüglich der Intention Z" (Stachowiak 1983: 118).

Im Zusammenhang mit Modellversuchen weist der Modellbegriff darauf hin, dass es sich um "[...] geistige Gebilde (handelt; Anm. Verf.), die durch planmäßige Konstruktionen auf der Basis perspektivischer Annahmen von Betrieb und Ausbildung, von Gesellschaft und vom Menschen gewonnen werden. In diesem Sinn lassen sich die Ergebnisse auch als kognitive Modelle bezeichnen" (Benteler 1991: 7). Es handelt sich demnach um reduzierte Abbilder, die ein gedankliches und dann im Modellversuch reales Probehandeln ermöglichen; Probehandeln in einem relativ geschützten Raum, der sich durch gute personelle und materielle Ausstattung auszeichnet.

Kritisch ist die Bezeichnung ‚Versuch‘ zu werten, da entgegen der Bedeutung von ‚Versuch‘ die Möglichkeit des Fehlschlagens weitestgehend ausgeschlossen wird. Die BLK fordert dazu auf, die Zielsetzungen so zu formulieren und die Versuchsbedingungen so zu beschaffen, dass nach aller Voraussicht ein positives Ergebnis erreicht wird.⁷

Ziel dieser Vorgehensweise ist es, die Betroffenen vor Misserfolgen zu schützen. Daraus folgt, dass der Begriff des ‚Versuchs‘ mit der Vorgabe in Form der ‚Erfolgsverpflichtung‘ korreliert, so dass dieser als missverständlich anzusehen ist.

Die begriffliche Klärung wirft demnach an dieser Stelle Unklarheiten auf: Wenn ein Scheitern grundsätzlich ausgeschlossen wird, ist es fraglich, ob es sich dann tatsächlich noch um ein ‚Erprobungsvorhaben‘ handelt. Es entsteht der Eindruck, dass vielmehr eine Durchsetzung politisch gewollter Entwicklungen gerechtfertigt werden soll und es sich dann bei Modellversuchen lediglich noch um Maßnahmen zur Entscheidungsbestätigung handelt.

Vor diesem Hintergrund scheint die Verwendung der Begrifflichkeit ‚Modellvorhaben‘⁸ angebrachter, da hierbei kein Versuchscharakter impliziert ist. Ob sich das Instrument den Vorwurf einer politischen Maßnahme zur Legitimation gewollter Entwicklungen gefallen lassen muss, wird die weitere Untersuchung zeigen.

7) vgl. dazu auch: BLK: Modellversuche in der Bewährung. Bonn 1995. S. 4; Schaade, K.: Modellversuche - eine Möglichkeit für die Weiterentwicklung beruflicher Bildungspraxis aus bildungspolitischer Sicht. In: Modellversuche - Ein Instrument zur Weiterentwicklung beruflicher Bildungspraxis. Hrsg. BiBB. Berlin: Schroedel Verlag 1978. S. 240.

8) Aus Gründen der einheitlichen Sprachverwendung wird auf den etablierten Begriff ‚Modellversuch‘ zurückgegriffen. Die damit einhergehenden Probleme bleiben jedoch bewusst.

2.1.2 Organisationsbezogene (administrative) Begriffserläuterung

Im Bereich der beruflichen Bildung werden Modellversuche vom Bundesministerium für Bildung und Forschung (BMBF) gefördert und an berufsbildenden Schulen sowie in der außerschulischen Berufsbildung, in Betrieben durchgeführt. Demnach werden schulische Modellversuche und Wirtschaftsmodellversuche auf Bundesebene unterschieden, während Schulversuche in alleiniger Länderhoheit, d. h. auf ein Bundesland begrenzt, initiiert werden.

Gemäß BerBiFG §6, Abs. 2, 1(d) fällt die Betreuung, Verwaltung und Förderung der Modellversuche in der außerschulischen Berufsbildung seit 1977 in die hoheitliche Obhut des Bundesinstituts für Berufsbildung, das diese Aufgaben auf Weisung des BMBF wahrnimmt (vgl. Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie: 1994). Hier werden Modellversuche als 'Versuchsmaßnahmen mit Erprobungscharakter' definiert (vgl. Bundesinstitut für Berufsbildung, zit. n. Euler 1995: 3).

Schulische Modellversuche dagegen werden in der Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung beraten, die ihren Auftrag aus dem Grundgesetz Artikel 91b erhält.

"Grundsätzlich dient die Modellversuchsförderung in der Berufsbildung dem Ziel staatlicher Berufsbildungspolitik, mit öffentlicher Finanzhilfe Leistung und Qualität des Berufsbildungssystems zu steigern, wobei die Berufsbildungspolitik mit anderen Politikbereichen in Verbindung steht, vornehmlich der Wirtschaftspolitik, der Arbeitsmarktpolitik und der allgemeinen Bildungspolitik" (Dehnbostel 1998: 188).

Es wird deutlich, dass zwar eine administrativ-bezogene Definition von Modellversuchen möglich ist, dass diese Instrumente, lt. Dehnbostel (ebd.), jedoch einen klaren Auftrag beinhalten und somit funktionalisiert sind. An dieser Stelle muss dann die Frage gestellt werden, für welchen Bereich, Politik oder Schule/Wirtschaft, der jeweilige Modellversuch Nutzen bringt, um so eine genauere Funktionsbestimmung vornehmen zu können.

2.1.3 Funktionale Begriffserläuterung

Wie oben schon deutlich gemacht, gibt es mehrere funktionale Erklärungsvarianten, die sich nach der Verwendung/dem Einsatzgebiet von Modellversuchen unterscheiden. Angesichts der zu untersuchenden 'Veränderungen' im Praxisfeld erfolgt eine Beschränkung der definitorischen Begriffsentwicklung, beziehungsweise auf diesen Anwendungsbereich:

"Modellversuche sind exemplarische Veränderungen in sozialen Feldern, wobei gleichzeitig berücksichtigt werden muss: Soziale Felder sind von Personen und Institutionen beherrscht und durch sie gestaltbare Ausschnitte der Erziehungswirklichkeit" (Sloane 1992: 10). Oder anders ausgedrückt:

Bei Modellversuchen handelt es sich "[...] um komplexe Entwicklungs- und Gestaltungsprojekte, die in einem durch Institutionen regulierten Praxisfeld entwickelt, er-

probt und evaluiert werden" (Euler; Sloane. „Implementation als Problem der Modellversuchsforschung“ <<http://www.wiso.uni-erlangen.de/WiSo/BWI/WiPd/seiten/virtpubl.html>>).

Die beiden Zitate machen deutlich, dass Modellversuche etwas mit beispielhaften Veränderungen im Rahmen eines begrenzten Bereiches (Praxisfeld, soziales Feld) zu tun haben, innerhalb dessen etwas erstmalig eingesetzt wird. Hervorzuheben ist dabei einerseits, dass institutionelle Hoheiten bestehen, die diese Felder reglementieren, andererseits, dass diese Felder begrenzt sind. Die Grenzen der Felder sind dabei durch den räumlichen und den zeitlichen Einsatzbereich der Projekte/ Modellversuche festgelegt.

Des Weiteren misst sich der Erfolg dieses Instruments u. a. an der Übertragbarkeit der 'beispielhaften Veränderungen' in den (Praxis-) Alltag (vgl. BLK 1995: 4); dieser Vorgang (Übertragung der Modellversuchsinhalte in die Modellversuchsschule und in andere Bildungseinrichtungen) wird als Implementation bzw. als Transfer bezeichnet (siehe hierzu auch Kap. 2.2.1).⁹

Durch die geplante Verbreitung der Modellversuchsergebnisse liegt eine doppelte Funktion von Modellversuchen darin, 'Abbild von etwas' und 'Vorbild für etwas' zu sein. Nach der Analyse des Untersuchungsgegenstandes 'Bildung' können Abbilder (Modelle) erstellt werden, die als vereinfachte Darstellung der Wirklichkeit (*Abbild von etwas*) angesehen und als Explikationsmodell oder Explikationsmodellversuch¹⁰ bezeichnet werden können. Entwickelte Modellversuche wiederum dienen als Vorbild für neue Experimente, sind also *Vorbild für etwas*, sog. Planungsmodelle (vgl. Benteler 1991: 8).

Eine funktionale Erklärungsvariante hebt demnach neben gegebenen 'Rahmenbedingungen' (Feldgrenzen) vor allem hervor, welcher Art der 'Nutzen', welcher Art die 'Leistungen' sind, die ein Modellversuch erbringen kann. In Bezug auf den Begriff der 'Veränderung' und in Anlehnung an die Ausführungen von Nickolaus, der sich mit der Erhöhung der Wirksamkeit innovatorischer Bildungskonzepte befasst, kann festgehalten werden:

Eine notwendige Bedingung für (nachhaltige) Verhaltensänderungen von Individuen im Praxisfeld (z. B. Schüler, Lehrer) stellt das Vermögen des Instruments Modellversuch dar, alternative, d. h. von der aktuellen (Bildungs-) Situation abweichende, Handlungsmöglichkeiten initiieren zu können. Solche Veränderungen werden dann als 'Innovation' bezeichnet, wenn diese, [...] zumindest bezogen auf Mo-

9) "Die Ergebnisse der Modellversuche sollen nicht nur für die jeweilige Schule, Hochschule oder andere Einrichtung von Bedeutung sein, sondern für das Bildungswesen des Landes und auch anderer Länder genutzt werden können, d. h. sie müssen übertragbar sein." (BLK 1995: 4) In Verbindung mit der Erfolgsverpflichtung stellt dies eine logische Konsequenz dar.

10) Der Begriff 'Explikation' vom lat. 'explicare' ist in diesem Zusammenhang nicht als 'Erklärung von etwas' zu verstehen, sondern vielmehr als 'Entfaltung/Erörterung von etwas/Auseinandersetzung mit etwas', nämlich (mit) der vereinfachten Darstellung eines Ausschnitts aus der Realität, dem Modell resp. Modellversuch.

dellversuchsort und –zeit und die daran Beteiligten, Neuigkeitswert besitzen. Im pädagogischen Feld können Veränderungen innovativen Charakters im Prinzip in allen Dimensionen des pädagogischen Prozesses auftreten" (Nickolaus 2001: 98)¹¹. Bezieht man an dieser Stelle die Forderung der BLK in die Betrachtung ein, dass die Ergebnisse aus Modellversuchen nicht nur (aber auch; Anm. Verf.) für die jeweilige Schule von Bedeutung sein sollen, sondern auch für das ganze Bildungswesen eines Landes (s. o.), dann wird deutlich, dass es sich bei diesen Ergebnissen um 'interne Effekte' - Leistungen für die jeweiligen Modellversuchsschulen (Implementation) und um externe Effekte - Leistungen für andere Bildungseinrichtungen (Transfer) handelt.

Aus den vorliegenden Ausführungen ergibt sich folgende für die Arbeit geltende Definition des Begriffes 'Modellversuch':

Modellversuche sind komplexe Projekte innerhalb vorher festgelegter räumlicher und zeitlicher Grenzen mit dem Ziel, die betrachteten Bereiche zu verändern und die Ergebnisse in die Modellversuchsorganisation (interne Effekte/Implementation) und nach außen zu übertragen (externe Effekte/Transfer).¹²

Für das Verständnis der nachfolgenden Ausführungen ist es außerdem von Bedeutung, die Tätigkeit im Modellversuchsfeld, die Modellversuchsarbeit (den Tätigkeiten in) der Berufsbildungsforschung bzw. die beiden Begrifflichkeiten direkt einander gegenüberzustellen, um festzustellen, ob Modellversuchsarbeit als Forschungstätigkeit anzusehen ist:

"Berufsbildungsforschung untersucht die Bedingungen, Abläufe und Folgen des Erwerbs fachlicher Qualifikationen sowie personaler und sozialer Einstellungen und Orientierungen, die für den Vollzug beruflich organisierter Arbeitsprozesse bedeutsam erscheinen" (DFG 1990: 1).

Eine Konkretisierung dessen, mit deutlichem Bezug zu den grundlegenden Wesensmerkmalen der Berufsbildungsforschung findet sich in der Formulierung:

"Für diese Berufsbildungsforschung ist konstitutiv, dass das in der Realität Vorgefundene innerhalb eines normativen pädagogischen Zielhorizonts dargestellt und mit Sollensvorstellungen konfrontiert wird" (DFG 1990: 19).

Bei der Gegenüberstellung der Begriffe 'Modellversuch' und 'Berufsbildungsforschung' werden durch den Vergleich (semantische) Parallelen und Differenzen deutlich:

Die Definition 'Modellversuch' ist deutlich Agens zugewandt, d. h. hier wird der Veränderungswille eines selektierten Feldes der Berufsbildung hervorgehoben ('Berei-

11) Zu der in dieser Arbeit zu untersuchenden 'Dimension pädagogischer Prozesse' siehe Kap. 4.1.

12) Zum Transferbegriff siehe Kap. 2.2.1.

che *verändern und transferieren*‘; s. o.). In Opposition hierzu ist der Begriff der Berufsbildungsforschung Agens abgewandt, was in der Verwendung des Verbs *‘untersucht’* zum Ausdruck kommt; gleiches gilt für die Konfrontation mit den in der Realität *‘vorgefundenen’* Aspekten.

Hieraus folgt, dass im Rahmen von Modellversuchsaktivitäten auf Erkenntnisse der Berufsbildungsforschung zurückgegriffen werden muss, um Teilaspekte der Berufsbildung zu verändern bzw. hierauf Einfluss nehmen zu können, was dem Gestaltungswillen der Modellversuche entspricht. Die Berufsbildungsforschung generiert hierfür die notwendigen theoriebezogenen Grundlagen. Demnach sind die Tätigkeiten und Erkenntnisse der Berufsbildungsforschung den Modellversuchstätigkeiten chronologisch und hinsichtlich ihres Selbstverständnisses vorgelagert. Darüber hinaus können Modellversuche selbst auch Gegenstand des Erkenntnisinteresses der Berufsbildungsforschung sein, um zu einer Weiterentwicklung dieser Erprobungsvorhaben Beiträge zu generieren.

In diesem Zusammenhang können Tätigkeiten im Modellversuchsfeld nicht mit Tätigkeiten der Berufsbildungsforschung gleichgesetzt werden. Lediglich die ex-post Betrachtung von Modellversuchen, ihre Analyse und Bewertung unter ausgewählten Gesichtspunkten (*‘Konfrontation mit Sollensvorstellungen’*; s. o.) kann unter den Forschungsbegriff subsumiert werden, da hier der primäre Gestaltungswillen von Berufsbildung nicht zu erkennen ist und die Untersuchungstätigkeit innerhalb des in der Definition des Begriffes Berufsbildungsforschung geforderten *‘normativen pädagogischen Zielhorizonts’* erfolgt, während Modellversuche ein jeweils spezielles Verwendungsinteresse, das durch die Transferorientierung zum Ausdruck kommt, inne haben.

2.2 Inhaltsbereiche der wissenschaftlichen Diskussion

Der Stand der wissenschaftlichen Diskussion, wie er in Publikationen aus Modellversuchen heraus, Veröffentlichungen der WBL von Modellversuchen, Schriften über einzelne Modellversuche sowie wissenschaftlichen Analysen nachzuvollziehen ist, bezieht sich auf die nachfolgenden Themenschwerpunkte im Bereich der Modellversuchsarbeit, die seit 1970 festzustellen sind:

Themenschwerpunkte Modellversuchsarbeit¹³

- Transferproblematik von Modellversuchsergebnissen
- Modellversuche und Organisationsentwicklung
- Leistungsfähigkeit von Modellversuchen - Modellversuche als Träger von Innovationen für die Berufsbildungspraxis/ Ergebnisse von Modellversuchen
- Methodologische Diskussion zur Begleitforschung

13) siehe hierzu u. a.: Dehnbostel 1992; Bader 1992, Höpke 1992; Dehnbostel 1995; Kutt 1995, Euler 1995; Bähr 1995; Dybowski 1995; Novak 1995; Zimmer 1995; Sloane 1992.

2.2.1 Transferproblematik von Modell Versuchsergebnissen

In den Veröffentlichungen der BLK zur Modellversuchsförderung wird hervorgehoben, dass die Ergebnisse der Modellversuche nicht nur für die jeweilige Schule, Hochschule oder andere Einrichtung von Bedeutung sein sollen, "[...] sondern für das Bildungswesen des Landes und auch anderer Länder genutzt werden (sollen; Anm. Verf.), d. h. sie müssen übertragbar sein" (BLK 1995: 4).

Aus diesen Anforderungen der Bund-Länder-Kommission ergeben sich unter Transfergesichtspunkten folgende Fragestellungen:

- Was sind 'Ergebnisse' von Modellversuchen?
- Wie ist der Transfer, das Übertragen von Ergebnissen, definiert?
- Findet ein Transfer der Modell Versuchsergebnisse realiter statt?

Euler vertritt die Ansicht, dass die Bandbreite der **Ergebnisse**, die im Rahmen von Modellversuchen erzielt und transferiert werden sollen, von der Unterstützung bei der Erfüllung gesetzlicher Mindestanforderungen bis zur aktiven Suche nach didaktisch-organisatorischen Prototypen der Berufsbildungspraxis reiche (vgl. Euler 1995: 227)¹⁴.

Dieses *ergebnisbezogene* Transferverständnis induziert zwei unterschiedliche Transferaspekte:

Einerseits ist ein produktorientierter Aspekt hervorzuheben, d. h. im Mittelpunkt der Verbreitung von Modell Versuchsergebnissen stehen 'übertragungsfähige Produkte' in Form von Materialien (siehe hierzu auch Benteler 1991: 10). Andererseits sind auch Merkmale eines Wandels hin zu einem prozessbezogenen Transferaspekt zu verzeichnen (vgl. Dehnhostel 1994: 6; Novak 1995: 63 ; Diener 1995: 292 u. a.), der seinen Ausdruck auch in einem entsprechenden Verständnis der Evaluationsaufgabe der WBL findet (vgl. Kap. 4) und bei dem Ergebnisse parallel zu ihrer Erarbeitung in andere Bildungsbereiche weitergegeben werden.

Über die Verbreitung von produkt- und prozessbezogenen Ergebnissen aus der Modell Versuchsebene hinaus, wird in dieser Arbeit allerdings auch die Formulierung der Erkenntnisinhalte aus der Evaluation des Modellversuchs, im Rahmen der wissenschaftlichen Begleitung als transferierbares Ergebnis angesehen, deren Güte mit den Leistungen eines Modellversuchs korrelieren (siehe Kap. 1.2).

In Bezug auf die somit angesprochenen Dokumentation der Erkenntnisse aus Modellversuchen ist relevant, dass die getroffenen Formulierungen, ein sprachlicher Ausdruck, ein Begriff, nicht identisch ist mit dem 'Phänomen', auf das sich sein 'Vorstellungsinhalt' bezieht, da er immer nur eine selektiv getroffene Auswahl von Bedeutungsinhalten repräsentiert, nie aber den ganzen Gegenstand (vgl.

14) Euler hebt außerdem hervor, dass Problemlösungen, die im Kontext von Modellversuchen entwickelt wurden, als Ergebnis von Modellversuchen gleichzusetzen seien mit Innovationen in Einzelbereichen einer heterogenen Berufsbildungspraxis. Insofern lasse sich nicht nur von Transfer im modellversuchsbezogenen Sinne, sondern auch von Innovationstransfer sprechen (siehe Euler 1995: 227).

Prim/Tilmann 1979: 38). D. h. die sprachlich formulierten Erkenntnisse der wissenschaftlichen Begleitung sind Konstruktionen der Wirklichkeit auf Grundlage subjektiver Erfahrungsstrukturen; diese durchlaufen einen mehrstufigen 'Transformationszusammenhang': Wahrnehmung von 'Sinnesdaten', geistige Verarbeitung und sprachliche Äußerung (vgl. Euler 1994: 218 u. Euler 1995: 231). Daraus folgt, dass der Übergang von einem Explikationsmodellversuch in einen Planungsmodellversuch einen Transformationsprozess erfordert, in dem die Gültigkeit und Anwendbarkeit der vorliegenden Erkenntnisse ein weiteres Mal geprüft werden müssen.

Neben diesen ergebnisbezogenen Aspekten von 'Transfer' ist außerdem eine Unterscheidung zwischen den unterschiedlichen Adressatengruppen, der Chronologie des Transfers sowie der Transferebene für die genauere Begriffsbestimmung notwendig:

Der Begriff der 'Implementation' bezieht sich auf die Einführung der Modellversuchsinhalte in den Modellversuchsprozess und damit auf die Konfrontation der Beteiligten und Betroffenen (Transfer im engeren Sinne; TeS) mit diesen Inhalten. Mit 'Transfer' dagegen ist die Verbreitung außerhalb der Modellversuchsorganisationen angesprochen (Transfer im weiteren Sinne; TwS). Da die Implementation (TeS) zunächst im Rahmen des je spezifischen Explikationsmodellversuches stattfindet, ist sie dem Transfer (TwS) zeitlich vorgelagert.

Darüber hinaus muss überlegt werden, auf welcher Ebene der Transfer stattfindet, d. h. kommt es auf Grund der Erkenntnisse aus einem Modellversuch zu einer Kodifizierung, z. B. in Curricula, dann kann von einer 'oberen Transferebene' gesprochen werden; führt dagegen ein Modellversuch bzw. die Erkenntnisse hieraus (nur) zu einer Intensivierung in der Zusammenarbeit von Schulen, entsteht z. B. ein Netzwerk, dann ist eine 'untere Transferebene'¹⁵ angesprochen. Da die BLK auf die Nutzung von Ergebnissen aus Modellversuchen auf Landes- bzw. Bundesebene abstellt (s. o.), ist der Transferbegriff im weiteren Sinne sowie der Transfer auf oberer Transferebene nicht nur für diese Arbeit, sondern auch generell von Relevanz:

Transfer meint die Verbreitung von produkt- und prozessorientierten Modellversuchsergebnissen außerhalb der jeweiligen Modellversuchsschule (TwS) und die Kodifizierung dieser Erkenntnisse in Richtlinien bzw. Rahmenrichtlinien oder anderen verbindlichen Vorgaben. Über die Ergebnisse aus der Modellversuchsebene hinaus ist die Verbreitung der evaluationsbezogenen Dokumentationen der wissenschaftlichen Begleitung (Evaluationsebene) unter den Begriff des Transfers zu subsumieren.

Um den **Erfolg des Transfers** sicher zu stellen, muss gewährleistet sein, dass die sachlich-intentionalen Bedingungen, d. h. die Frage, ob die Probleme der Ziel-

15) Nickolaus (2001a) spricht in diesem Zusammenhang von 'institutioneller und individueller Handlungsebene'.

organisation mit denen, für die die Problemlösungen entwickelt wurden, gleichzusetzen sind. In einem zweiten Schritt muss überlegt werden, inwieweit in der Zielorganisation die notwendigen Handlungskompetenzen wie Problembewusstsein, innovatives Denken und Veränderungsbereitschaft bei denjenigen vorhanden sind, die diese Innovationen umsetzen sollen. Und nicht zuletzt sind es institutionelle Bedingungen, wie materielle und kulturelle Voraussetzungen, die über den Transfererfolg entscheiden. Hinzu kommen Probleme, wie die Verbreitung der Modellversuchsergebnisse, die Frage des Zeitpunktes des Transfers u. v. a., die Einfluss auf das Verbreitungsergebnis nehmen (vgl. Euler, Kutt 1995: 271).

Bei dieser Vielzahl von (additiv aneinander gereihten) Einflussvariablen stellt sich die Gefahr ein, dass Modellversuchsergebnisse häufig nicht oder nur zum Teil transferiert werden (können). Neben den genannten sachlich-intentionalen sind es möglicherweise modellversuchsübergreifende Aspekte, wie sie analog z. B. bei der Übernahme wissenschaftlicher Erkenntnisse in Anwendungsbereiche auftreten, die die Adaption von Modellversuchsergebnissen erschweren oder verhindern.

Euler greift in seinen Ausführungen 'Denn sie tun nicht, was sie wissen' diesen Problembereich auf (Euler 1996: 350): In dem o. g. Aufsatz geht er der Frage nach, welche Gründe zu einer begrenzten Berücksichtigung wissenschaftlicher Theorien in der Praxis führen, um im Analogieschluss zu er(be-)gründen, warum dokumentierte Modellversuchsergebnisse nicht in vergleichbaren Praxisfeldern angewandt werden. Zur Beantwortung zieht er zwei Argumentationslinien heran:

Einerseits sei die Praxis aus der wissenschaftlichen Theorie heraus nicht vollständig erreichbar. Es handele sich um träges Wissen, das erst durch die Fähigkeiten der Anwender praxiswirksam werde. Wissenschaftliche Theorien seien Abstraktionen, eine Reduktion gegebener Realität. Doch gerade die als konstant angesehenen Randbedingungen könnten in der Praxis zu bedeutsamen Variablen werden, die es unbedingt zu berücksichtigen gelte.

Andererseits forciere die Vielzahl vorhandener Theoriegebäude die Schwierigkeit der Anwendung vorliegender Ergebnisse. Gerade für die Berufs- und Wirtschaftspädagogik gelte, dass es keine ‚herrschende Meinung‘ gebe, d. h. aus dem vorliegenden Theoriebestand müsse also die ‚passende‘ ausgewählt und eine Bewertung der zum Teil widersprüchlichen Aussagen vorgenommen werden (Euler 1996: 354).

Es wird deutlich, dass neben den sachbezogenen Problemen (sachlich-intentional, personell, institutionell) auch modellversuchunspezifische, d. h. adaptionsbezogene Erschwernisse bei der Übertragung wissenschaftlich fundierter Erkenntnisse bestehen, die das Transfervorhaben verzögern oder verhindern können. Eine Lösung dieses Dilemmas wird in der Literatur nicht aufgezeigt (siehe auch BMBF 2001 und Kap. 2.3.1). Empirische Untersuchungen, die sich auf die obere Transferebene beziehen, fehlen. Es liegen lediglich Ergebnisse aus einer Umfrage unter 'Akteuren' (Mitarbeiter in der wissenschaftlichen Begleitung zu MV) vor, die

auf die Selbsteinschätzungen dieser Personengruppe in Bezug auf die jeweiligen Modellversuche abstellt (BMBF 2001: 82ff; siehe auch Nickolaus/Schnurpel 2001b).

Das Fehlen empirischer Transferuntersuchungen kann als Beleg herangezogen werden, um Aussagen zur Nachhaltigkeit von Modellversuchen in Frage zu stellen; Angaben über Innovationseffekte von Modellversuchen sind demnach (siehe Nickolaus/Schnurpel 2001b) lediglich nicht belegte Vermutungen, was die Dringlichkeit dieser Arbeit über die 'Leistungsfähigkeit' von Modellversuchen unterstreicht.

2.2.2 Modellversuche und Organisationsentwicklung

Seit einigen Jahren wird im Rahmen der Modellversuchsarbeit auch der Zusammenhang von Modellversuchen und Organisationsentwicklung diskutiert. In diesem Kontext werden Modellversuche als Instrumente beschrieben, "[...] die in soziale Systeme eingreifen, Innovationen in der beruflichen Bildung anstreben und daher stets Veränderungsprozesse auslösen und Veränderungen bewirken" (Kutt 1995: 241; siehe Kap. 2.1.3 ≈ Funktionen von MV).

Die Aspekte 'Innovation anstreben' und 'Veränderungen bewirken' beziehen sich auf Produkt-, Verfahrens- und Prozessinnovationen, die transferiert werden und somit ihre singuläre Bedeutung im je spezifischen Modellversuch aufgeben. Der prozessorientierte Transferbegriff (siehe Kap. 2.2.1) macht deutlich, dass durch die wechselseitige Rückkoppelung zwischen den am Modellversuch beteiligten Schulpraktikern und den Mitarbeitern der WBL Veränderungsprozesse initiiert werden; gemeint sind 'Verfahrens- und Einstellungsänderungen' sowie Umgestaltungen der organisatorischen, technologischen und ökonomischen Strukturen, wodurch der Bezug zur Organisationsentwicklung hergestellt wird.

Eine mögliche Funktion von Modellversuchen als Instrumente der Organisationsentwicklung (OE) soll durch den Vergleich der Ziel-Inhalts-Dimensionen beider Ansätze nachgegangen werden:

Die Gesellschaft für Organisationsentwicklung (GOE) versteht unter OE einen "[...] längerfristigen, organisationsumfassenden Entwicklungs- und Veränderungsprozess von Organisationen und der in ihr tätigen Menschen. Der Prozess beruht auf Lernen aller Betroffenen durch direkte Mitwirkung und praktische Erfahrung. Sein Ziel besteht in einer gleichzeitigen Verbesserung der Leistungsfähigkeit der Organisation (Effektivität) und der Qualität des Arbeitslebens (Humanität)" (GOE 1983, zit. n. Kieser 2001: 119).

Daneben gibt es auch andere, von dieser Definition abweichende Begriffsbeschreibungen, aus denen Kieser gemeinsame Merkmale herausgearbeitet hat und die für einen Vergleich mit den aus der Definition 'Modellversuch' erkennbaren Intentionen herangezogen werden können (s. u.).

Da es sich bei einer Definition, wie die hier für Modellversuche geltende, im Gegensatz zu einer Sammlung von Merkmalen um eine Verdichtung von Informatio-

nen handelt, in die verschiedene Gestaltungsmerkmale einfließen, müssen die bei der Definitionsfindung bedeutsam gewordenen, aber nicht explizit genannten Merkmale mit herangezogen, d. h. hervorgehoben werden, damit sie in einem solchen Untersuchungsschritt wirksam werden können. Es handelt sich neben den Grundcharakteristika von Modellversuchen (siehe Def. von MV) um Aspekte der institutionellen Hoheit/administrativen Vorgaben sowie der relevanten Evaluationsstandards (s. o.)¹⁶.

In Bezug auf die Berücksichtigung der Evaluationsstandards bei diesem Vergleich der Ziel-Inhalts-Dimensionen ist folgender (Exkurs) Zusammenhang für das grundlegende Verständnis von Bedeutung:

Ein Aufgabenbereich von Modellversuchen stellt die wissenschaftliche Begleitung dar, deren Dokumentationen sich als Teil der (Berufs-) Bildungsforschung manifestieren (siehe Kap. 1.2). Hierin enthalten sind u. a. die Erkenntnisse aus der Analyse und der Bewertung des Modellversuchsaufbaus und der einzelnen Modellversuchsaktivitäten, die als dokumentierte Ergebnisse der Evaluation angesehen werden können. Diese Ergebnisse und die entsprechenden Evaluationsaktivitäten haben sich an Qualitätsstandards zu orientieren, die somit eine Gestaltungsfunktion erhalten und die Modellversuchsaktivitäten prägen. Neben dem summativen Bereich der Gesamtevaluation kommt es durch den formativen Teil zu einem wechselseitigen Rückkoppelungsprozess von Evaluationsstandards resp. Evaluationstätigkeiten und dem Modellversuchsprozess.

Die Subsumtion der von Kieser herausgearbeiteten Merkmale unter die Definition des Instruments Modellversuch (siehe Kap. 2.1.3) lässt die abschließende Aussage zu, ob Modellversuche als Instrumente der Organisationsentwicklung anzusehen sind:

- **Prozessorientierung, d. h. im Vordergrund steht die Entwicklung der Organisation vom Ist-Zustand (Z_0) zum Soll-Zustand (Z_1)**

Der Begriff der ‚Veränderung‘, wie er auch für das Instrument Modellversuch verwendet wird, hebt den der Organisationsentwicklung innewohnenden Prozesscharakter hervor. Dieser Aspekt wird einerseits durch die Implementation des Modellversuchs in die Modellversuchsinstitution (Schule) sowie durch den Transfer in andere Einrichtungen berücksichtigt (siehe auch Kap. 2.2.1 'wechselseitige Rückkopplung').

- **Unterstützung des Veränderungsprozesses durch die Betroffenen (Betroffene zu Beteiligten machen)**

Ein Qualitätsstandard von Modellversuchsevaluationen ist die Ermittlung der Beteiligten und Betroffenen, damit deren Interessen und Bedürfnisse berücksichtigt wer-

16) Als Qualitätsstandards gelten hier die Standards des Joint Committee on Standards for Educational Evaluation als Anforderungen an die Evaluation eines Modellversuchs (siehe hierzu Kap. 1.2 und Kap. 5).

den können und somit die Motivation gegeben ist, den Veränderungsprozess zu unterstützen.

▪ **Mitwirkung eines intervenierenden Beraters**

Von der BLK wird in §7 der ‚Rahmenvereinbarung zur koordinierten Vorbereitung und Durchführung und wissenschaftlichen Begleitung von Modellversuchen im Bildungswesen‘ verpflichtend die Mitwirkung der wissenschaftlichen Begleitung bei der Durchführung (und Evaluation) eines Modellversuchs festgelegt (vgl. BLK 1976: §7).

Das Selbstverständnis der Mitglieder der WBL, das in der Begleitung des Modellversuchs zum Tragen kommt, hat sich seit Initiierung von Modellversuchen 1971 gewandelt. Auch wenn in den verschiedenen Modellversuchen unterschiedliche methodische Herangehensweisen verwendet werden, ist generell eine Entwicklung hin zu einem 'Beratermodell' der WBL festzustellen (siehe hierzu Dehnbostel, Demuth 1995: 46f u. Kap. 4).

▪ **Zielkomplementarität – Ziele der Organisation und Ziele der Organisationsmitglieder können und sollen gleichzeitig verfolgt werden**

Auch hier gelten die Hinweise für den Aspekt 'Unterstützung des Veränderungsprozesses durch die Betroffenen (Betroffene zu Beteiligten machen)' in Bezug auf die Funktion der WBL, dessen Berücksichtigung durch die Einhaltung des Evaluationsstandards N1 (Ermittlung der Beteiligten und Betroffenen) sicher gestellt ist.

Darüber hinaus hebt der Evaluationsstandard 'Unterstützung der Dienstleistungsorientierung/K1' hervor, dass die Evaluation so geplant werden muss, dass die Interessen und Bedürfnisse des ganzen Zielgruppenspektrums zu berücksichtigen sind.

Die vorstehenden Ausführungen zeigen Parallelen/Übereinstimmungen, die als Indizien angesehen werden können, um das Instrument Modellversuch auch als Maßnahme der Organisationsentwicklung anzusehen.

Jedoch sind 'Maßnahmen der Organisationsentwicklung' und 'Modellversuche' in unterschiedlichen Entstehungszusammenhängen verwurzelt und werden im Kontext unterschiedlicher sozialer Systeme verwandt, deren Zielausrichtung weitestgehend verschieden ist – "Die Ansätze entstammen unterschiedlichen Traditionen: Während Modellversuche ursprünglich als Instrument einer rationalen Bildungsplanung geschaffen wurden [...], entstanden die Ansätze der OE in der betriebswirtschaftlichen Organisationslehre [...]" (Euler 1995: 280).

In der Tendenz ist festzustellen, dass sich Organisationsentwicklung vornehmlich in Wirtschaftsorganisationen 'vollzieht', die durch ein hohes Maß an ‚Ökonomischer Rationalität‘ gekennzeichnet sind, während Modellversuche überwiegend in Organisationseinheiten des Bildungswesens durchgeführt werden, die nach Ansicht von Euler im Spannungsfeld von 'Ökonomischer Zweckrationalität' und 'Pädagogischer Sinnrationalität' stehen, so dass Innovationen schon wegen der gegebenen rechtlichen und hoheitlichen Rahmenbedingungen nicht in direkt vergleichbarem Maß und Weg durchgeführt werden können.

Gemein hingegen ist beiden Ansätzen das Ziel einer Entwicklung und Erprobung zielorientierter Innovationen in Organisationen, wobei Probleme der Praxis Anlässe von Modellversuchen und Organisationsentwicklung sind. Weiterhin werden die Innovationen als Problemlösungsprozess verstanden, der durch eine Einbeziehung der Betroffenen sowie durch Prozessunterstützung durch Berater, d. h. Organisationsberater im Rahmen der OE und die WBL im Fall der Modellversuche, gekennzeichnet ist (vgl. Euler 1995: 280), wobei das Ausmaß der Zielkomplementarität in Bezug auf die Ziele der Organisationsleitung und den anderen Mitgliedern der Organisation unterschiedlich sein kann.

Neben dem Hervorheben von Gemeinsamkeiten tragen auch Akzentuierungsunterschiede, die sich auf den ‚typischen Fall‘ beziehen, zu einer weiteren Abgrenzung der Begrifflichkeiten bei:

So kann festgehalten werden, dass Modellversuche tendenziell ergebnisorientiert, während OE-Ansätze in der konzeptionellen Ausrichtung eher prozessorientiert sind. Dies ist u. U. darauf zurück zu führen, dass Modellversuche grundsätzlich zeitlich begrenzt sind und ihr Erfolg an den transferfähigen Ergebnissen gemessen wird. Weiterhin müssen Modellversuche der pädagogischen Sinnrationalität Rechnung tragen. Organisationsentwicklung betont dagegen eher die Bedeutung anderer Subsysteme, wie das der Leistungserstellung (vgl. Euler 1995: 281).

Grundsätzliche Kritik, Modellversuchsforschung als Organisationsentwicklung anzusehen, stammt von Dehnbostel. Diese begründet er mit der Gleichstellung von Praktikern und Forschern in einigen handlungstheoretischen Paradigmen, die zu einer Reduzierung wissenschaftlicher Begleitung auf die Funktion des Beraters und Organisationsentwicklers führt. Die nötige Distanz zum Modellversuchsfeld vermindere die Möglichkeit der gesicherten wissenschaftlichen Erkenntnisgewinnung, da der handlungstheoretische Diskurs von Wissenschaft und Praxis einseitig auf eine praxis- und organisationsgeleitete 'Subjekt-Subjekt-Ebene' verlagert werde.

"Unabhängige wissenschaftliche Standards und methodologische Grundlagen sind Voraussetzung dafür, das Verhältnis von Wissenschaft und Praxis als einen wechselseitigen Lernprozess zu gestalten, zumal dann, wenn Beobachtung und Beschreibung im Forschungsprozess¹⁷ in starkem Maße in Abhängigkeit von der subjektiven Erfassung der Wirklichkeit zu sehen sind" (Dehnbostel: 1998: 192f).

2.2.3 Leistungen von Modellversuchen - Modellversuche als Träger von Innovationen für die Berufsbildungspraxis

"Die Modellversuche sollen so ausgerichtet sein, dass sie wichtige Entscheidungshilfen für die Entwicklung des Bildungswesens geben" (BLK 1976: 31). Oder konkreter formuliert, "sie dienen der Entwicklung und Erprobung neuer inhaltlicher, struktureller und organisatorischer Ansätze im Bildungswesen. Die Modellversuche sollen die Grundlage für Aussagen zu bildungsplanerischen und bildungspolitischen

17) Ob der Begriff der 'Forschung' im Zusammenhang mit der Tätigkeit im Modellversuchsfeld gerechtfertigt ist, soll in Kap. 4.1 erörtert werden.

Fragestellungen verbessern und einen Beitrag zur Weiterentwicklung des Bildungswesens leisten" (BLK 1981: 7). Das Bundesinstitut für Berufsbildung (BiBB) formuliert ähnlich, jedoch schon deutlich abgeschwächt, was die Leistungen von Modellversuchen angeht. Hiernach sollen Innovationen 'mittlerer Reichweite' erreicht werden (BiBB 1997: Vorwort).

Für diese auf 'Nachhaltigkeit' ausgerichteten Erprobungsvorhaben wurden allein 1997, dem letzten Jahr des Untersuchungszeitraumes, vom Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie 19,78 Millionen € für die Durchführung und Evaluation im schulischen und betrieblichen Bereich zur Verfügung gestellt¹⁸, um der staatlichen Berufsbildungspolitik zu dienen und um mit öffentlicher Finanzhilfe die Leistungskraft und Qualität des Berufsbildungssystems zu steigern. Hierbei ist man von der Annahme ausgegangen, dass ein hohes Qualifikationsniveau die Voraussetzung für die wirtschaftlich-technische und ökologische Entwicklung darstellt und die internationale Konkurrenzfähigkeit sowie eine zukunftsgerichtete Flexibilität des Arbeitsmarktes sichert.

Darüber hinaus wird in der Literatur im Zusammenhang mit der 'Leistungsfähigkeit' von Modellversuchen auch ihre Funktion als Instrument zur Demokratisierung von Entwicklungsprozessen im Bildungswesen, als partizipationsorientiertes, innovatives Mittel hervorgehoben (Dehnbostel u. a. 1994: 4f).

Wie oben schon erläutert (vgl. 2.1.3), werden Modellversuche in der Regel (so die Forderung der BLK) an ihren transferfähigen Ergebnissen gemessen. Eine erste Einschätzung ihrer Leistungsfähigkeit auf Grundlage der 'Erfahrungen' der ersten zehn Berichtsjahre erfolgte 1981 durch die BLK, die in ihrer Festschrift '10 Jahre Modellversuche im Bildungswesen' zu dem Ergebnis kommt, "[...] dass Modellversuche in aller Regel weit weniger spektakuläre Ergebnisse erbringen, als dies ursprünglich wohl von den Beteiligten und der Öffentlichkeit erwartet worden ist" (BLK 1981: 6).

In den Veröffentlichungen der BLK oder des BiBB wird dieses Problem jedoch in den Jahren darauf nicht wieder aufgegriffen. Es wird nicht diskutiert oder auch nur darauf hingewiesen, dass die Leistungserwartungen nicht erfüllt wurden.

Erst 1993 erscheint wieder ein Bericht über die Auswertung von 28 Modellversuchen, die in der Zeit von 1980 bis 1989 zum Themenbereich 'Neue Informations- und Kommunikationstechniken in der beruflichen Bildung' durchgeführt und von der BLK gefördert wurden. Faber und Kaiser, die die Untersuchung leiteten, entwickelten zunächst einen Fragebogen, der von den verschiedenen Modellversuchs-Arbeitsgruppen beantwortet und mit den Zwischen- und Abschlussberichten zurückgesandt wurde und somit als Basis für die Auswertung dienen sollte. Sie erstellten eine Übersicht über Inhalte und Ergebnisse der Modellversuche, so dass die Auswertung den Charakter eines Nachschlagewerkes erhalten sollte. In einem

18) Diese Daten entsprechen den Mittelzuweisungen des BMBF an die BLK und das BiBB und sind von diesen beiden Institutionen für die vorliegende Studie zusammengestellt worden.

weiteren Schritt untersuchten sie Rahmenrichtlinien und Ausbildungsordnungen auf Neuerungen, die auf die durchgeführten Modellversuche zurückzuführen sind.

Leider ist die gesamte Vorgehensweise bei dieser Untersuchung nur schwer nachzuvollziehen; insbesondere bleibt offen, wie der Fragebogen aufgebaut und welche Informationen durch diesen erhalten wurden. Auch die Analyse der Rahmenrichtlinien und Ausbildungsordnungen lässt sich nicht nachvollziehen. Gleiches gilt daher auch für das nachfolgende Ergebnis der Untersuchung, das besagt,:

"[...] dass die Modellversuche zu vielfältigen Innovationen in den Bereichen Curriculumdesign und Lehrplanentwicklung, didaktische Konzepte und Unterrichtsmethodik, Fachraumkonzepte und Medienentwicklung, Ausstattung mit Hardware und Software sowie Lehrerfortbildung und Lehrerweiterbildung geführt haben" (BLK 1993: 4).

Auf einer Fachtagung 1994, anlässlich des 20-jährigen Bestehens des Bundesinstituts für Berufsbildung widmete man sich wiederum diesem Themenkomplex und stellte die Frage: 'Was leisten Modellversuche?' Ein Jahr später, 1995, erschien die Zusammenfassung der Themen und Beratungsergebnisse, die gleichsam als 'Leistungsschau der Modellversuche' titulierte wird. Es war beabsichtigt, die Ergebnisse der acht Arbeitsgruppen darzustellen, die einen 'lebendigen und repräsentativen Querschnitt der Modellversuchsthemen' widerspiegeln sollten.

Tatsächlich wird die Ausgangsfrage nicht beantwortet, vielmehr werden Ergebnisse, Strukturen und Konzepte vorgetragen und beraten (siehe Bähr, Holz 1995: 17). Eine nachvollziehbare und anhand von Kriterien durchgeführte Leistungsprüfung wird auch hier nicht dokumentiert oder diskutiert.

Anlässlich des 1997 stattfindenden Forums zur Berufsbildungsforschung in Nürnberg ging man im Rahmen eines Workshops der Frage nach: 'Berufliches Lernen als Forschungsgegenstand – Inwieweit sind Modellversuche in der Berufsbildung ein Instrument von Erkenntnisgewinnung und Praxisinnovation?' Die Frage verdeutlicht die beiden potentiellen Schwerpunkte des Workshops.

Von den insgesamt zehn verschriftlichten Beiträgen zu diesem Workshop beziehen sich jedoch nur zwei auf den Bereich Praxisinnovation und lediglich ein Beitrag zeigt die Veränderungen der Berufsbildungspraxis nach Durchführung eines Modellversuches auf. Die verbleibenden Veröffentlichungen stellen auf grundsätzliche Fragestellungen der Modellversuchsforschung ab; genauer, inwieweit Modellversuche in der Berufsbildung zu neuen Erkenntnissen für die Berufsbildungsforschung führen können. Man verfolgte Fragestellungen, wie die nach dem anzuwendenden Forschungsansatz bzw. der wissenschaftstheoretisch-methodologischen Position der Begleitforschung und es wurden Fragen aufgeworfen, ob privatwirtschaftlich organisierte oder universitär gebundene Institutionen die wissenschaftliche Begleitung übernehmen sollen. Gleichwohl kommen die Veranstalter in ihrem Resümee zu dem Ergebnis, dass Erkenntnisgewinnung für die Modellversuchsarbeit wohl konstitutive Bedeutung einnimmt, "[...] dass Forschung (jedoch; Anm. Verf.) nicht der Anlass, sondern entsprechend der von der BLK getroffenen Bestimmung von Modellversuchen eine ‚Zutat‘ von Modellversuchen ist" (Euler, Dehnhostel 1998: 495). Anders ausgedrückt ließe sich hieraus die Aufforderung formulieren, dass die

Begleitforschung von Modellversuchen nicht das Ziel verfolgen sollte, vorrangig Leistungen für die Berufsbildungsforschung zu erbringen, sondern vielmehr ist die Objektebene, gemeint ist hier der einzelne Modellversuch, zu analysieren und die entstandenen Effekte sind herauszustellen und zu bewerten.

2.3 Erste Zwischenbilanz: Ausprägungen der inhaltlichen Schwerpunkte in der Modellversuchsliteratur¹⁹

Transferproblematik:

Die zum Teil vorhandenen additiven Auflistungen von transferbezogenen 'Erfolgsindikatoren' zeigen (vermutlich) nicht die relevanten Indikatoren auf oder diese werden nicht ausreichend berücksichtigt.

Modellversuche und Organisationsentwicklung:

Da die berufsbildende Schule als komplexes 'System' anzusehen ist, das sich von marktwirtschaftlichen Unternehmen unterscheidet, können Modellversuche, die sich nur auf Teilbereiche des Systems beziehen, auch nur als 'partielle Organisationsentwicklungsansätze'²⁰ angesehen werden.

Auseinandersetzung mit den 'Leistungen' von Modellversuchen:

Die Beschäftigung mit den Bereichen für die Modellversuche Leistungen erbringen können erfolgt in der Literatur nicht ausgewogen. Im Hinblick auf das Praxisfeld 'Schule' werden 'innovative Rekonstruktionen', die die Weiterentwicklung und Verbesserung schulischer Arbeit in den Mittelpunkt stellen, nicht oder nicht ausreichend erörtert.

2.3.1 Auseinandersetzung mit der Transferproblematik von Modellversuchsergebnissen

Im Rahmen der Auseinandersetzung mit diesem Themenbereich werden vor allem 'sachlich-intentionale' Bedingungen hervorgehoben, die bedeutend für den Transfer von Modellversuchsergebnissen sind. Die genannten Faktoren unterscheiden sich in Abhängigkeit vom Autor der jeweiligen Publikation hinsichtlich ihrer Quantität und der Formulierung; inhaltliche Überschneidungen sind jedoch festzustellen (siehe hierzu Rottluff 1992: 265ff; Benteler 1991 u. a.).

Höpke und Wolf fassen diese transferbezogenen erfolgsrelevanten Aspekte in den Punkten 'materialinterne und materialexterne Faktoren' zusammen und schaffen damit eine Möglichkeit der Systematisierung der in diesem Kontext bedeutsamen Bedingungen (vgl. Höpke, Wolf 1992: 257).

Zu den materialinternen Faktoren zählen die Relevanz des Produktes für potentielle Anwender, die gewonnene Qualität, Deutlichkeit und Komplexität (vgl. Fullan 1983: 491). Bei den materialexternen Faktoren dagegen ist zu unterscheiden zwischen

19) Unter Modellversuchsliteratur werden hier Dokumentationen verstanden, die Publikationen aus Modellversuchen heraus, Veröffentlichungen der WBL von MV, Schriften über einzelne MV sowie wissenschaftliche Analysen zu diesem Themenbereich darstellen.

20) Siehe zum Vergleich die Definition von 'Organisationsentwicklung' der GOE, Kap. 2.2.2.

den "[...] objektiven betrieblichen Bedingungen des 'übernehmenden Betriebs' im Unterschied zum Modellversuchsbetrieb, z. B. Betriebsgröße, Technikausstattung, Arbeitsorganisation, Produktpalette oder Qualifikationsstand, die Orientierung und Grundauffassung der Mitarbeiter, aber auch die personalen Bedingungen, die Rolle der Geschäftsleitung und des Ausbildungsleiters in der betrieblichen Bildungsarbeit, die Interaktion zwischen den Ausbildern, ihr Verständnis von betrieblicher Bildungsarbeit und von dem Verhältnis Ausbilder – Auszubildender. Weitere material-externe Faktoren sind die Zugänglichkeit der Informationen über die Modellversuche und die Möglichkeiten einer Interaktion zwischen dem Modellversuchsträger und den Betrieben außerhalb des Modellversuchs" (Höpke, Wolf 1992: 257).

Die Autoren, die sich mit den genannten Erfolgsindikatoren auseinandersetzen, verstehen diese als Antezedenzbedingungen, die die Entwicklung vom Explikations- zum Planungsmodellversuch und die Kodifizierung der Modellversuchsergebnisse begünstigen oder aber auch behindern können.

Die Feststellung der BLK im Rahmen ihrer Festschrift zum zehnjährigen Bestehen von Modellversuchen im Bildungswesen, dass Modellversuche weit weniger leisten als ursprünglich erwartet (BLK 1981: 6), lässt in Bezug auf die Transferproblematik die Vermutung zu, dass 'Erfolgsindikatoren' in der Modellversuchsarbeit entweder nicht (ausreichend) berücksichtigt werden oder dass es nicht oder nicht allein die hervorgehobenen Indikatoren sind, die für den Transfer relevant sind.

Euler dagegen ordnet die Transferfrage in den Zusammenhang der Frage nach den Bedingungen der Anwendungen von (wissenschaftlichen) Theorien ein, da keine expliziten Untersuchungen über die Anwendung von Modellversuchsergebnissen bekannt sind. Dabei stellt er nicht die Modellversuchsergebnisse mit Theorien gleich, sondern nutzt diese Herangehensweise für eine Analogiebetrachtung, in welcher Form wissenschaftliche Theorien in der jeweils relevanten Praxis aufgenommen und verwendet werden. Anders als in seinem Aufsatz 'Denn sie tun nicht, was sie wissen' (siehe Euler 1996) geht er hierbei nicht nur von den beiden genannten Begründungslinien (Praxis ist aus der wiss. Theorie heraus nicht vollständig erreichbar und Vielzahl vorhandener Theorien erschweren die Anwendung von Erkenntnissen; siehe auch Kap. 2.2.1) aus, sondern erörtert die Transferfrage in erkenntnis-, transfer- und organisationsentwicklungstheoretischer Perspektive (vgl. Euler 1995: 225ff) und kommt dann zu Aussagen mit präskriptivem Charakter, die den Transfer von Modellversuchsergebnissen erleichtern sollen. Durch diese Form der Analogiebetrachtung begibt Euler sich mit seiner Herangehensweise auf eine Metaebene und weicht damit von der Aneinanderreihung 'relevanter' Transfermerkmale ab, um so die Nachhaltigkeit des Instruments Modellversuch steigern zu können:

▪ Erkenntnistheoretische Perspektive:

"Die Übertragung von Problemlösungen [...] aus Modellversuchen wird im Verständnis kognitiv-konstruktivistischer Erkenntnistheorien nicht als ein Kopier-, sondern als ein Auswahl- und Konstruktionsprozess – und damit strenggenommen als ein neuer Problemlösungsprozess – gedacht. Demnach werden Problemlösungs-

angebote [...] aus Modellversuchen von einem potentiellen Anwender auf der Grundlage seiner jeweils aktuell wirksamen Alltagstheorien ausgelegt [...]" (Euler 1995: 237).

▪ **Transfertheoretische Perspektive:**

"Problemlösungen [...] aus Modellversuchen dürfen in der Darstellung nicht auf ihren abstrakten Kern reduziert werden, sondern wichtig erscheint die nachvollziehbare Verbindung von generalisierbaren Aussagen und konkretisierenden Beispielen [...]" (Euler 1995: 243).

▪ **Organisationsentwicklungstheoretische Perspektive:**

"Problemlösungen [...] aus Modellversuchen determinieren nicht den Transfer, sondern sie sind auf die spezifischen Bedingungen der Zielorganisation hin auszulegen und anzupassen. Diese Offenheit kann im Extrem dazu führen, dass der Transfer sich von den Zielen und Maßnahmen im Ausgangskonzept löst und zu völlig entgegengesetzten Maßnahmen führt" (Euler 1995: 247).

Euler kommt zu dem Schluss, dass auch in der Wissenschaft die Anwendung wissenschaftlicher Theorien nicht friktionslos erfolgt. Vielmehr habe sich vor allem die erkenntnistheoretische Sicht bestätigt, wonach die Adaption von (wissenschaftlichen) Theorien nicht als technologische Anwendung nomologischer Aussagen, sondern als 'interessengesteuerte Integration' in Alltagstheorien verlaufe (vgl. Euler 1995: 252). In Analogie zu seinen auf die Adaption von 'Theorien' bezogenen Ausführungen wendet er sich dem Themenfeld 'Transfer von Modellversuchen' zu und hebt hervor, dass die Anwendung von Modell Versuchsergebnissen maßgeblich durch die organisatorisch-kulturellen Bedingungen des Praxisfeldes, hier der Schule, beeinflusst werde.

Die Produkte aus Modellversuchen sind demnach nicht handlungsleitend, sondern bedürfen der Aufbereitung, der Anpassung an das 'jeweilige Praxisfeld Schule' durch den 'jeweiligen Anwender', der mit der ihm eigenen Relevanzeinschätzung über die Aufnahme der 'jeweiligen Aspekte' aus dem Modellversuch entscheidet.

Die hervorgehobenen Transferbedingungen (materialintern und materialextern, s. o.) können somit nicht als additive Aneinanderreihung von Indikatoren angesehen werden, bei denen mit zunehmender Anzahl der Berücksichtigung der Transfererfolg zunimmt. Vielmehr hängt die Entscheidung hinsichtlich der Wertigkeit dieser Indikatoren vom adaptierenden Subjekt und der von ihm vorgenommenen Selektion der Indikatoren und deren Relevanzeinschätzung ab.

Will man den Transfer von Modell Versuchsergebnissen steigern, so erscheint es von Bedeutung, diesen Selektions- und Verbindungsprozess über die Anwendung von Modell Versuchsergebnissen explizit, z. B. in eigenständigen Modellversuchen, zu untersuchen und auf eine zunehmende Auflistung von 'Erfolgsindikatoren' zu verzichten.

2.3.2 Auseinandersetzung mit dem Aspekt ‚Modellversuche und Organisationsentwicklung‘

"Organisationsentwicklung erhebt den Anspruch, Strukturen, Personen, Zusammenhänge und Veränderungen in Organisationen ganzheitlich zu betrachten, während Modellversuche durch den Bezug auf Einzelprobleme und den Bezug auf den Berufsbildungsbereich zumeist nur Teilbereiche von Organisationen in Innovationsbereiche einbeziehen" (Dehnbostel 1995: 237).

Die von Dehnbostel hervorgehobenen Unterschiede zwischen Organisationsentwicklung und Modellversuchen, die vor allem durch das Hervorheben der 'Einzelfallbetrachtung' in Modellversuchen zum Ausdruck kommen, erfahren durch den in dieser Arbeit eingegrenzten Untersuchungsbereich 'schulischer Modellversuche mit dem Ziel der Erprobung von Lehr-, Lernmethoden' eine weitere 'Verschärfung', da über die Einzelfallbetrachtung hinaus eine selektive Betrachtung ausgewählter Aspekte in den Mittelpunkt gerückt wird.

Des Weiteren stellt die Organisation 'berufsbildende Schule als Systemeinheit' u. a. durch die rechtlichen Bedingungen (Landesschulgesetze und Beamtenstatus der Mitarbeiter), aber auch durch die Vielzahl verschiedenartiger Schulformen und Berufsfelder an vollständig unterschiedlichen Schulstandorten ein komplexes System dar, das sich von den rechtlich-organisatorischen Rahmenbedingungen marktwirtschaftlicher Unternehmen stark unterscheidet, so dass Modellversuche schon aus diesen Gründen nur als 'partielle' Organisationsentwicklungsansätze angesehen werden können. Ergebnisse aus Modellversuchen, die in einzelnen Schulformen der berufsbildenden Schule durchgeführt werden, können demnach nicht friktionslos auf andere Berufsfelder der gleichen (oder einer anderen) berufsbildenden Schule oder auf den Teilzeit- resp. Vollzeitbereich übertragen werden um der 'Systementwicklung' zu dienen, da die Profile der unterschiedlichen Berufsfelder (z. B. Wirtschaft/Verwaltung und Metall) bzw. Organisationsformen (Teilzeit-, Vollzeitbereich) nicht deckungsgleich sind.

"Grenzen sind (demnach; Anm. Verf.) sowohl durch das Ziel- und Orientierungssystem der auf Berufsbildung bezogenen Modellversuche, als auch durch die Komplexität von Gesamtorganisationsbetrachtungen gesetzt" (Dehnbostel 1995: 237).

2.3.3 Auseinandersetzung mit dem Aspekt der 'Leistungen' von Modellversuchen

Die Ausführungen in Kap. 2.2.3 zeigen, dass die Würdigung des Instruments Modellversuch unter dem Aspekt der 'Leistungen', die sie erbringen sollen, in den unterschiedlichen Arten von Literatur einen großen Raum einnimmt, dass jedoch klare Präferenzen zugunsten methodologischer und forschungspolitischer Aspekte gesetzt werden. Hierbei handelt es sich häufig um eine Auflistung von Postulaten, deren Hintergründe nicht ausführlich geprüft oder diskutiert werden. Die oben dargestellten Bereiche, für die Modellversuche Leistungen erbringen können (Bildungspolitik und -planung, Berufsbildungsforschung und Berufsbildungspraxis), sind in der Literatur nicht gleichmäßig vertreten, es wird ihnen keine ausgewogene Zuwendung zuteil.

In Bezug auf die Veränderungen des Praxisfeldes heben Kaiser und Weitz, deren Aussagen beispielhaft zur Verdeutlichung herangezogen werden, in ihrem Endbericht zum Modellversuch 'Neue Informationstechnologien und Datenverarbeitung im Berufsfeld Wirtschaft und Verwaltung – Verbindung von berufspraktischer und theoretischer Arbeit (Ausbildung) im Lernbüro' hervor, dass die wissenschaftliche Begleitung den Implementationsprozess dokumentieren soll, Analysen und innovative Rekonstruktionen zu leisten hat. "Unter innovativer Rekonstruktion versteht die WBL (wissenschaftliche Begleitung; Anm. Verf.) die Dokumentation der vorgefundenen schulischen Praxis verbunden mit dem Versuch, die zukünftige schulische Arbeit im Hinblick auf Weiterentwicklung und Verbesserung zu erörtern und zu unterstützen" (Kaiser, Weitz: 1990: 26).

Doch die angekündigte Erörterung verbesserungswürdiger Aspekte unterbleibt. Vielmehr erfolgt eine Dokumentation des Modellversuchverlaufes im Sinne einer Beschreibung der Situation an den beteiligten Schulen. Die fehlende Auseinandersetzung mit Problembereichen lässt den Modellversuch beliebig erscheinen, da keine Aussagen zur Wirksamkeit getroffen werden und somit eine weitere Verwendung der Ergebnisse in Frage gestellt ist. Damit ist dann auch die Weiterentwicklung dieses Modellversuches im Sinne einer Wendung vom Explikationsmodellversuch zu einem (weiteren) Planungsmodellversuch verbaut worden, so dass der finanzielle und personelle Aufwand nicht gerechtfertigt erscheint.

Zu kritisieren ist weiterhin die Entwicklung und nachfolgende Anwendung des Evaluationskonzepts. Nach Aussagen von Kaiser und Weitz ist dies nach dem Prinzip der qualitativen Unterrichtsforschung, der Handlungsforschung, ausgerichtet und soll als Entscheidungs- und Implementationshilfe dienen. Dabei sind es vor allem kommunikationsoffene Verfahren, die als Bewertungshilfe von Lernprozessen herangezogen werden: der Diskurs als elementares Prinzip dieser Evaluation. Hinzugezogen werden außerdem strukturierte Beobachtungen, qualitative Interviews und "[...]Erhebung und Analyse von Dokumenten unterschiedlichster Natur" (Kaiser, Weitz 1990: 25).

Für die weitere Argumentation und Vorstellung der Ergebnisse im Rahmen des Abschlussberichtes werden aber dann nur noch die Ergebnisse der Abschlussbefragung der beteiligten Schüler mittels Fragebogen herangezogen, bzw. die Ergebnisse der Befragung der beteiligten Lehrer genutzt (Kaiser, Weitz 1990: 67ff). Die Auswertung dieser Befragung dient nach eigenen Aussagen der Modellversuchsleitung, der Legitimation des Bildungsganges (Kaiser, Weitz 1990: 31). Bei diesem Vorgehen erscheint es zweifelhaft, ob der Gegenstand der wissenschaftlichen Begleitung, die innovative Rekonstruktion, nicht lediglich ein Postulat geblieben ist, denn eine Überprüfung der verfolgten Modellversuchsziele (und das in einer summarischen Evaluation festgestellte 'Maß der Erreichung evaluationsbezogener Ziele'), die Aussagen über die erreichten Veränderungen hervorbringen könnte, unterbleibt. Die Ziele werden nur in den Anträgen zum durchgeführten Erprobungsvorhaben aufgezeigt, nicht in den vorliegenden Berichten und sie werden nicht zur ex-post Beurteilung herangezogen. Hier drängt sich der Eindruck auf, dass eine Evaluierung der Vorgehensweise lediglich im Hinblick auf die Rechtfertigung (politisch) gewollter Innovationen vorgenommen wird.

3 Bildungspolitischer Rückblick und Untersuchung der finanziellen Förderungsentwicklung (Teil A der Studie)

3.1 Bildungspolitischer Rückblick

3.1.1 Ausgangslage

Eine reine chronologische Dokumentation von Bildungspolitik, bei der eine Einbettung des Instruments Modellversuch in die (Berufs-) Bildungspolitik der Bundesrepublik Deutschland erfolgt, würde keine Aussagen dahingehend zulassen, inwiefern Modellversuche als "[...] konstitutive Bestandteile von Bildungsplanung und Bildungsreform, die der Rationalisierung bildungspolitischer Entscheidungen dienen sollen" (Weishaupt 1992:19) genutzt wurden.

Daher ist es zweckmäßig, eine Untersuchung der Schwerpunkte und dem Wandel der Berufsbildungspolitik anhand eines Vergleiches von Berufsbildungszielen und deren Erreichung, die in den Berufsbildungsberichten dokumentiert wurde, durchzuführen und die initiierten Modellversuche diesen Ergebnissen gegenüber zu stellen, um zu überprüfen, ob Modellversuche initiiert wurden, die gemäß der o. g. Intention die bildungspolitische Zielrichtung der politisch Verantwortlichen unterstützt haben. Im vorliegenden Fall stehen die Legislaturperioden 1982 (1983) – 1998 im Mittelpunkt der Betrachtung; für die Zeit ab 1983 liegt dementsprechend auch eine Auflistung der geförderten Modellversuche seitens der BLK vor. Der Zeitraum vor 1983 ist in dieser Arbeit nicht erfasst worden, da die entsprechenden Unterlagen bei der BLK nicht mehr vorhanden sind.

3.1.2 Systematisierung der geförderten Modellversuche

Für die genannten Untersuchungsschritte ist es jedoch zunächst von Bedeutung, die durchgeführten Modellversuche zu systematisieren, da die BLK lediglich Förderbereiche definiert hat, wie z. B. den Bezugsbereich beruflicher Bildung, die in sich eine zu große Zahl unterschiedlich fokussierter Modellversuche beinhalten, als dass sie für eine Untersuchung der bildungspolitischen Zielrichtungen herangezogen werden könnten. Die vorhandene und von der BLK vorgenommene Differenzierung des Förderbereichs 'berufliche Bildung' in die Innovationsbereiche 'Modernisierung der Lerninhalte und Ausbildungsmethoden', 'Entwicklung adäquater Konzepte für die Berufsbildung besonderer Gruppen' und 'Transfer von Berufsbildungs-Innovationen in die neuen Länder' kann für das Vorhaben nicht genutzt werden, da diese das o. g. Problem der unterschiedlich intendierten Modellversuche auf Grund zu weit gefasster Gruppierungen nicht zu lösen vermag.

Auf Grundlage der 128 von der BLK aufgelisteten und im Verbund mit den Ländern geförderten Modellversuche wird anhand der genannten Modellversuchstitel die nachfolgende Systematisierung erstellt. Sollte dieses Vorgehen im Einzelfall zu Ungenauigkeiten führen, da die Titel nicht aussagefähig sind, so werden diese unklar betitelten Modellversuche nicht erfasst. Ausgeschlossen bleiben weiterhin die Modellversuche, die in den 'neuen' Bundesländern durchgeführt wurden, da zu er-

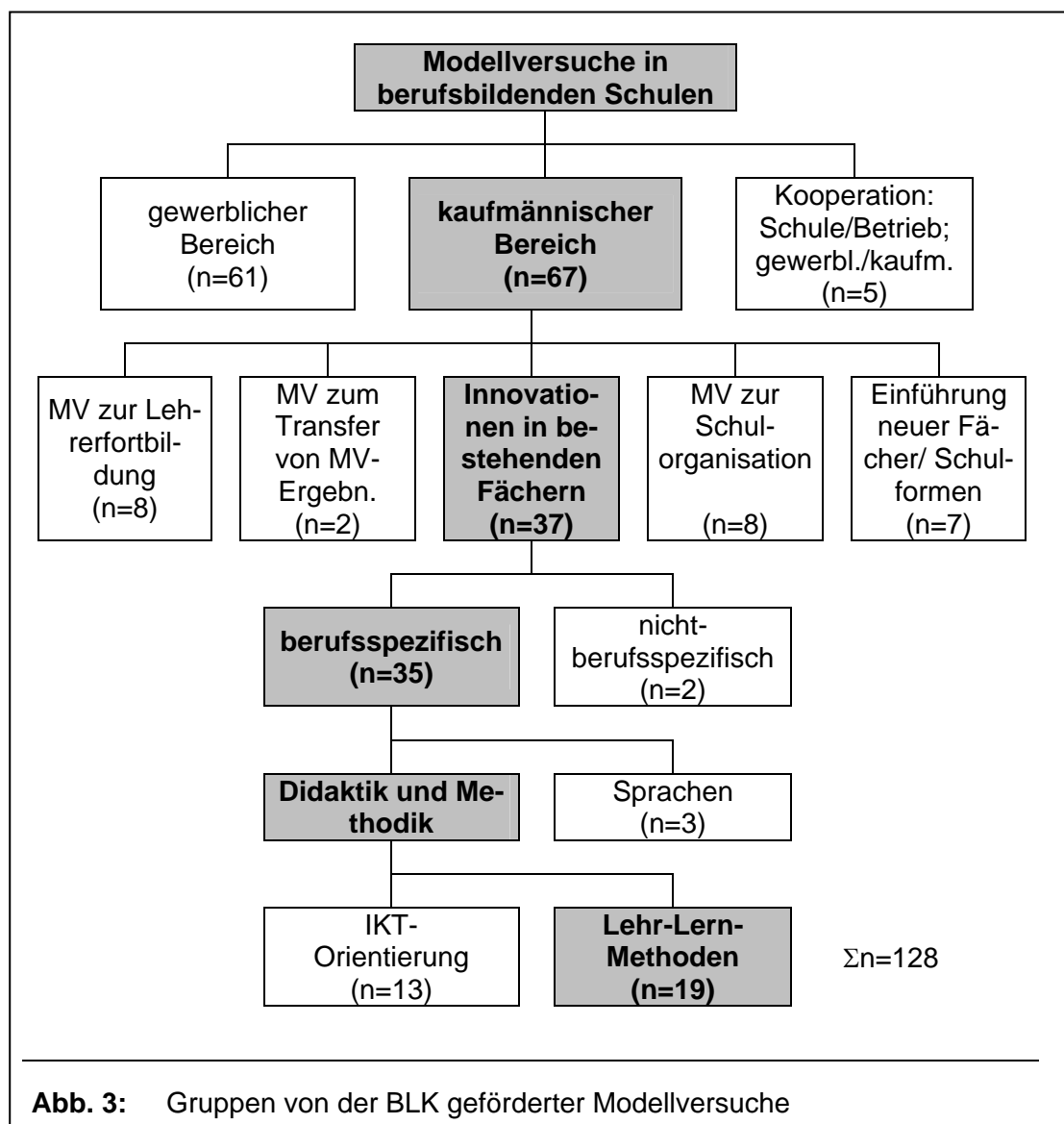
warten ist, dass hier andere politische Problembereiche im Untersuchungszeitraum zur Lösung vordringlich waren, die für den Bundesschnitt nicht repräsentativ waren und somit zu einer Verfälschung führen könnten.

Gruppe 1:	Modellversuche zur Kooperation der Lernorte Die Abstimmung/Kooperation der Partner im dualen System steht hier im Vordergrund. Neben dem Abbau von Hemmschwellen sollen auch neue Kooperationsformen vorbereitet werden.
Ziel:	Die Modellversuche sollen Kooperationsformen ermitteln und Möglichkeiten der Zusammenarbeit der Lernorte konstruieren.
Gruppe 2:	Modellversuche zur Lehrerfortbildung, Praktika, Tagungen und der Erstellung von Unterlagen und Curricula Diese Gruppe impliziert u. a. Fortbildungskonzepte, die Lehrer auf neue berufspraktische Anforderungen vorbereiten sollen. Aber auch die Vorbereitung auf schüleraktive Lernkonzepte fällt in diesen Bereich.
Ziel:	Die Modellversuche sollen Möglichkeiten der Lehrerbildung im tertiären Bereich aufzeigen und schüleraktive Lernkonzepte entwickeln.
Gruppe 3:	Transfermodellversuche Hierunter werden diejenigen Modellversuche gefasst, die sich mit dem Transfer von Modell Versuchsergebnissen in andere Bundesländer befassen bzw. sich um die Lösung der Implementationsproblematik bemühen.
Ziel:	Die Modellversuche sollen vorhandene Modell Versuchsergebnisse transferieren bzw. Lösungsansätze zum Problem der Implementation gestalten.
Gruppe 4:	Modellversuche zu Innovationen in bestehenden Fächern oder Schulformen Dieser Bereich zielt auf die Entwicklung neuer Lehr-, Lernmethoden sowie neuer Lehr-, Lernhilfen. Wegen der Vielzahl der Modellversuche, die diesem Bereich zuzuordnen sind, ist eine genauere Differenzierung notwendig. Zunächst sollen diejenigen MV aus der Betrachtung ausgenommen werden, bei denen ein Berufsbezug nicht direkt gegeben ist (Bsp. Sportunterricht). Auch die sprachlich orientierten Vorhaben sind gesondert auszuweisen. Die jetzt noch verbleibenden MV verfolgen didaktisch-methodische Ziele, wobei unterschieden werden kann zwischen MV, die sich mit der lehr-lernorganisatorischen Umsetzung von informations- und kommunikationstechnologischen Problemstellungen (Einsatz neuer Software u. a.) beschäftigen und solchen, die sich mit der Ausbil-

Ziel:	<p>derung der noch verbleibenden Handlungskompetenzen auseinanderzusetzen.</p> <p>Die berufsspezifisch intendierten Modellversuche dieses Bereiches sollen didaktisch-methodische Konzepte zur Förderung beruflicher Handlungskompetenz entwickeln. Dabei beziehen sie sich einerseits auf die Nutzung von Informations- und Kommunikationstechnologien, andererseits sind alternative Lehr-Lernorganisationsformen zu erproben und vorzustellen.</p>
Gruppe 5:	<p>Modellversuche zur Schulorganisation</p> <p>Dieser Bereich umfasst die Themen Organisationsentwicklung, Aufbau- und Ablauforganisation an beruflichen Schulen. Ziele dieser MV können demnach einerseits die Veränderung der internen Schulorganisation, andererseits auch der Außenbeziehungen sein. Beispielhaft sei hierfür auf internationale Kontakte oder die Einbindung verschiedener kultureller Gruppen in das soziale Feld ‚Schule‘ verwiesen.</p>
Ziel:	<p>Die Modellversuche sollen Beiträge zur 'Organisationsentwicklung' an beruflichen Schulen hervorbringen.</p>
Gruppe 6:	<p>Modellversuche zur Einführung neuer Fächer, Lerngebiete oder Schulformen</p> <p>Hierzu zählt sowohl die Entwicklung doppelqualifizierender Bildungsgänge als auch die Initiierung neuer Schulformen.</p>
Ziel:	<p>Die Modellversuche sollen neue Bildungsgänge definieren und ihre Konstruktion und Durchführung kritisch würdigen.</p>

Die nachstehende Abbildung gibt die Systematisierung der von der BLK im Zeitraum 1983 bis 1998 geförderten Modellversuche wieder. Da bei der vorliegenden Verteilung die Gruppe ‚Innovation in bestehenden Fächern‘ überproportional vertreten ist, könnte dies als Indiz dafür gewertet werden, dass hier ein besonderer Entwicklungsbedarf aus Sicht der verantwortlichen Bildungspolitiker vermutet werden kann.

Daher können die Ergebnisse dieser Systematisierung nicht nur für die beabsichtigte 'Ziel-Inhalts-Analyse' im Rahmen eines (berufs-) bildungspolitischen Rückblicks genutzt werden, sondern diese Modellversuchsgruppe wird außerdem aufgegriffen und im Rahmen einer metaevaluativen Untersuchung mit Hilfe eines exmanenten Rasters analysiert, um Aussagen zu den Leistungen der Erprobungsvorhaben tätigen zu können (siehe Kap. 6).



3.2 Ziel-Inhalts-Analyse von Berufsbildungsberichten in ihrer Funktion bildungspolitischer ex-post Dokumentationen

Aussagen zur Gestaltung und zum Wandel von Berufsbildungspolitik können durch die Analyse von Zielsetzung und Zielerreichung getroffen werden, während ein Vergleich der beiden Zielgrößen Schlussfolgerungen im Hinblick auf den (dargestellten) Erfolg bildungspolitischer Bemühungen zulassen. Dazu ist es notwendig, Zielformulierungen zu explizieren und hieraus homogene Gruppen zu bilden; anschließend sind diese zu operationalisieren und als immanentes Messraster an ex-post Dokumentationen wie dem Berufsbildungsbericht anzulehnen.

Die Explikation der Berufsbildungsziele kann anhand von Erklärungen der jeweiligen Bundesbildungsminister zu den Aufgabenschwerpunkten ihrer Amtszeit erfolgen, die in den offiziellen Veröffentlichungsorganen des Ministeriums oder der verschiedenen Bundesregierungen abgedruckt wurden. Hierzu dienen im Folgenden die 'Informationen Bildung und Wissenschaft' als Organ des Bundesbildungsminis-

teriums und andererseits das 'Bulletin', das die Arbeit der Bundesregierung wiedergeben soll.

Dabei hängt die Qualität der zu gewinnenden Erkenntnisse im Wesentlichen von der Ausdrucksstärke der Berufsbildungsziele ab. Sind diese lediglich derart formuliert, dass sie Absichtserklärungen darstellen, kann eine Operationalisierung nicht erfolgen und es sind lediglich Trendaussagen/Dokumentationen möglich, die eine exakte Formulierung des Zielerreichungsgrades vermissen lassen.

Doch auch in diesem Fall oder wenn Bundesbildungsminister vollständig auf die Dokumentation der zu verfolgenden Berufsbildungsstruktur verzichtet haben, sind Indizien über die Qualität der Berufsbildungspolitik gegeben.

Zunächst ist es erforderlich, die einzelnen Amtsperioden der fünf Bundesminister im Untersuchungszeitraum anhand der Zielformulierungen und der Aussagen aus den Berufsbildungsberichten zu dokumentieren (siehe Anhang zu Teil A). In Kap. 3.3.1 erfolgt lediglich eine Auswertung der Ergebnisse und ein Nachzeichnen der inhaltlichen Entwicklungslinien der Berufsbildungspolitik, damit initiierte Modellversuche in Kap. 3.3.2 in die Berufsbildungspolitik eingeordnet und Aussagen zur Kohärenz der bildungspolitischen Bemühungen vorgenommen werden können.

3.3 Auswertung der Untersuchung berufsbildungspolitischer Intentionen 1983 - 1998 – Zusammenfassung und Einordnung der Modellversuchsarbeit in den berufsbildungspolitischen Horizont

3.3.1 Entwicklungslinien in der Berufsbildungspolitik

Die Anwendung des immanenten Messrasters, d. h. die Explizierung von Berufsbildungszielen und deren Anlehnung an die ex-post Dokumentation Berufsbildungsbericht brachte keine Ergebnisse hervor, die Aussagen über den Zielerreichungsgrad zulassen. Die von den jeweiligen Bundesbildungsministern formulierten Ziele waren nicht geeignet derart operationalisiert zu werden, als dass sie Aussagen im Sinne von 'Ziel erreicht/Ziel nicht erreicht' bzw. prozentuale Annäherung an das gesetzte Ziel zuließen.

Dennoch ist der angewendete Untersuchungsschritt nicht obsolet, vielmehr lässt sich hierdurch die Frage rechtfertigen, wie berufsbildungspolitische Schwerpunkte Bedeutung gewinnen können und wie die zuständigen Minister bzw. deren Mitarbeiter erkennen, wann Arbeitsschwerpunkte erfüllt und neuere Entwicklungen in das berufsbildungsplanerische Zielkonzept aufgenommen werden müssen.

Rolff geht davon aus, dass "in allen Fällen [...] Bildungsplanung gewissermaßen auf den Vollzug bereits abgeschlossener Willensbildungsprozesse beschränkt (ist; Anm. Verf.). Planung hält sich dabei im Rahmen der technischen Rationalisierung längst erfolgter sozialer Prozesse und politischer Entscheidungen" (Rolff 1971: 319).

Hierfür nennt er drei Gründe: Zum einen wird die Schere zwischen gesellschaftlichen Entwicklungen und dem Bildungswesen so groß, dass die Notwendigkeit von Veränderungen²¹ offensichtlich ist. Oder, zweitens, der politische Druck auf die Betroffenen wird so stark und findet in der Öffentlichkeit derart Unterstützung, dass die Entscheidungsträger nachgeben müssen. Zuletzt können auch Forschungsergebnisse den öffentlichen Willensbildungsprozess derart beeinflussen, dass die politischen Entscheidungsträger zum Handeln gezwungen werden.

Daraus folgt, dass Bildungsplanung als Bestandteil der Bildungspolitik häufig auf die Umsetzung eines abgeschlossenen Willensbildungsprozesses beschränkt ist und als rein 'technische Realisierung' charakterisiert werden kann (vgl. Rolff 1971: 319f).

Demnach sind Zielformulierungen des jeweiligen Bundesministers, die auf ein selbständiges Agieren abzielen, (zunächst einmal) nicht erforderlich. Diese können aus den Forderungen relevanter gesellschaftlicher Bereiche entnommen und zu einem berufs-bildungspolitischen Programm zusammengefasst werden. Dabei soll nicht die Leistung geschmälert werden, die in der Abstimmung mit den anderen Politikbereichen, wie Wirtschafts-, Finanz- und Arbeitsmarktpolitik, erforderlich ist. Auch die Umsetzung der Forderungen oder Zielformulierungen zu Umsetzungsplänen stellt erhebliche Anforderungen.

Dennoch ist das Messen bzw. Prüfen der Zielerreichung nicht zu unterlassen; die Ergebnisse hieraus sind, ungeachtet der oben gemachten Ausführungen, notwendig, damit Ressourcen zielgerichtet eingesetzt werden können. Spätestens an dieser Stelle sind operationalisierte Zielformulierungen unabdingbar und bei Nachlässigkeiten Kritik an dem Vorgehen der zuständigen Bundesminister für den Untersuchungszeitraum angebracht.

Bislang (ebenfalls) unbeachtet ist der Aspekt der prospektiven Initiierung berufsbildungspolitischer Entwicklungen, um arbeitsmarktpolitischen Entwicklungen, auch Fehlentwicklungen, vorzugreifen und eigenständig zu agieren, statt lediglich auf fremdbestimmte Entwicklungen zu reagieren. Fehlt dieses selbständige Handeln, kann nicht von einer geplanten bildungspolitischen Entwicklung, in der Modellversuche ihre Funktion als Instrument einer auf gesicherten Erkenntnisse basierenden Politik wahrnehmen, gesprochen werden.

Zusammenfassend kann zum Problem der mangelnden Zielformulierung in Bezug auf die vorliegende Untersuchung festgehalten werden:

Nichtoperationalisierbare Zielformulierungen verhindern die eigene Zielkontrolle und somit eine Beurteilung der geleisteten Arbeit. Auch eine öffentliche und sachbezogene Kommunikation über den Ressourceneinsatz wird verhindert. Nicht zuletzt hat die mangelnde Zielformulierung negativen Einfluss auf eine aktive Berufsbildungspolitik, da lediglich auf gesellschaftliche Entwicklungen reagiert werden kann. Darüber hinaus wird auch eine Kontinuität der geleisteten Arbeit verhindert,

21) Der Begriff der 'Reform' wird von mir an dieser Stelle bewusst nicht verwendet, da diese Argumentation auch auf den regelmäßigen Bildungsplanungsprozess zutrifft.

denn nachfolgende Verantwortliche können nicht über die notwendigen weiteren Tätigkeiten informiert werden.

Die Intentionen der Berufsbildungspolitik lassen sich auf Grund der gewonnenen Erkenntnisse in (wenige) Schwerpunktbereiche unterteilen und zusammenfassen. Diese wurden innerhalb der Amtszeiten der Bildungsminister mit unterschiedlicher Intensität verfolgt. Die folgende Übersicht bietet einen Überblick über die oben ausformulierten Politikbereiche und deren Umsetzung in den jeweiligen Amtsperioden der Bundesminister:

Berufsbildungspolitische Schwerpunkte	Willms	Möllemann	Ortleb	Laermann	Rüttgers
Stärkung des Ausbildungsplatzangebotes	X	X	X		X
Veränderung von Rechtsvorschriften (Aus- u. Fortbildungsordnungen)	X	X		X	
Problemgruppen (Jugendliche o. Schulabschluss, Ausländer, Frauen u. a.)	X	X			
Weiterbildung	X	X			
Zusammenarbeit Bund und Länder	X				
Neue Technologien	X	X			
Durchlässigkeit und Gleichwertigkeit	X	X	X	X	
Differenzierung und Begabtenförderung	X		X	X	
Europa			X	X	X
Aufbau Ost			X		
Attraktivitätssteigerung des dualen Systems ²²			X	X	X

Tab. 1: Bildungspolitische Schwerpunkte

Mit Ausnahme der kurzen Amtsperiode des Bundesbildungsministers Laermann ist die Stärkung des Ausbildungsplatzangebotes durchgängig als wichtiger berufsbildungspolitischer Schwerpunkt angesehen worden.

Parallelen sind vor allem bei Willms und Möllemann und anschließend bei Ortleb (Laermann) und Rüttgers zu erkennen. Zwischen diesen Parallelgruppierungen gibt es jedoch einen offensichtlichen Bruch in der Kontinuität der Intentionen. Der Anlass hierfür geht aus den Zielformulierungen selbst nicht hervor. Jedoch scheint das Bildungsverhalten der Jugendlichen die Ursache für den berufsbildungspolitischen Richtungswechsel zu sein. Denn 1992, also mit der Übernahme des Bundesbildungsministeriums durch Ortleb, ist erstmals ein Rückgang der Neuabschlüsse im Bereich der Berufsausbildungsverträge zu verzeichnen (BMBF 1993: 3). Im gleichen Zeitraum hat die Zahl der Studienanfänger erheblich zugenommen. Diese

22) Diese Formulierung ist insofern ungenau, als dass sie mehrere Maßnahmen zur Attraktivitätssteigerung beinhaltet, wie Durchlässigkeit, Gleichwertigkeit usw., die oben auch genannt sind. Da dieser Aspekt jedoch neben den anderen Punkten hervorgehoben wurde, ist er in die Auflistung aufgenommen worden.

Situation spitzt sich in den Folgejahren zu, was nicht nur an den absoluten Zahlen der Studienanfänger abzulesen ist; vor allem wird die Berufsausbildung als Sprungbrett für berufliche Aufstiegschancen angesehen, wenn hieran ein Studium angeschlossen wird (BMBF 1993 ff). Dies führt zu erheblichen Schwierigkeiten der Industrie, des Handels vor allem aber des Handwerks, da der Nachwuchsbedarf nicht gedeckt werden kann.

Der Hauptausschuss des Bundesinstituts für Berufsbildung weist in diesem Zusammenhang auch auf die Aussichtslosigkeit verschiedener Ausbildungsberufe hin. Fehlende Aufstiegschancen und schlechte Ausbildungsqualität sind den Bewerbern bekannt und führen vor allem zu einer verstärkten Nachfrage in aussichtsreicheren Berufen (BMBF 1993: 20).

Während also im Zeitraum 1983 bis etwa 1991 vor allem der Jugendliche selbst im Mittelpunkt der Betrachtung stand und viel Initiative in den Förderbereich 'benachteiligte Jugendliche' investiert wurde, galt hiernach das Schlagwort 'Attraktivitätssteigerung des dualen Systems' als wichtigste berufsbildungspolitische Intention.

Berücksichtigt man in diesem Zusammenhang noch einmal die o. g. Aussagen zum Problem der Zielformulierungen, dann wird deutlich, dass die Berufsbildungspolitik zumindest ansatzweise auf gesellschaftliche oder wirtschaftliche Entwicklungen lediglich reagiert. Dies wird vor allem ersichtlich zu dem Zeitpunkt, als Rüttgers 1994 das Ministerium von seinem Amtsvorgänger Ortleb (Laermann) übernahm; er unterließ vollständig eine Dokumentation und Veröffentlichung der von ihm verfolgten berufsbildungspolitischen Grundstrukturen. Die Beweggründe für sein Handeln bleiben damit offen.

3.3.2 Kohärenz der Modellversuchsarbeit mit der Berufsbildungspolitik des Untersuchungszeitraumes

Die nachfolgende Übersicht stellt eine Zusammenfassung der in dem Untersuchungszeitraum initiierten Modellversuche dar. Es erfolgt eine Zuordnung der Modellversuche zu Ziel-Gruppen²³ und den Jahren, in denen die jeweiligen Vorhaben gestartet wurden. Hierdurch wird ein Überblick möglich, der, wenn er den berufsbildungspolitischen Intentionen gegenübergestellt wird, Auskunft über den Zusammenhang von Modellversuchen und Berufsbildungspolitik geben kann. Es erfolgt eine Konzentration auf die 'Gruppe 4b – Modellversuche zu didaktisch-methodischen Entwicklungen (Lehr-, Lernmethoden)', die als Gruppe mit der stärksten Ausprägung festgestellt wurde (siehe Kap. 3.1.2). Die hierin identifizierten Zielbereiche werden expliziert, um die Transparenz des Vorgehens zu steigern.

23) Die Ziel-Gruppen entsprechen der in Kap. 3.1.2 vorgenommenen Systematisierung.

Gruppe/Jahr	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
Gruppe 1: Kooperation			1						1	1			1	1		
Gruppe 2: Lehrerfortbildung		2		1		1			1			1	2			
Gruppe 3: Transfer									2							
Gruppe 4a: Innovation - IKT- Orientierung		4	1		1	1				1		2		1	2	
Gruppe 4b: Innovation did.-meth.		4					2		4	1	1	1	3	1	2	
Gruppe 4c: Innovation –Sprachen											1		1			
Gruppe 5: Schulorganisation	1	1				1						1	1	1	1	
Gruppe 6: Neue Fächer/ Schulformen		2	1	1				1				1	1			

Tab. 2: Initiierungsjahre ausgewählter Modellversuche

Des Weiteren können die Aussagen zur Kohärenz Modellversuchsarbeit – Berufsbildungspolitik für die Amtstätigkeiten der Minister Willms und Möllemann bzw. Ortleb (Laermann) und Rüttgers zusammengefasst werden, da, wie oben dargestellt, die berufsbildungspolitischen Intentionen Parallelen aufzeigen; auf diesem Weg wird eine weitere Konkretisierung des Untersuchungsblickwinkels erreicht:²⁴

Die in der Amtszeit der Ministerin Willms (1983 - 1987) durchgeführten Modellversuche der Gruppe 4b beschäftigen sich mit dem Aspekt der praxisnahen Unterrichtsgestaltung, der konzeptionellen Verbesserung des Unterrichts an Berufsfachschulen und der Teilzeitberufsschule sowie der Entwicklung eines pädagogisch-didaktischen Konzeptes für Teilzeitberufsschulen von Jugendlichen, die besonderer Lernhilfen bedürfen.

Die in der Amtszeit Möllemanns (1988 bis 1991) initiierten Modellversuche im Geltungsbereich der Gruppe G 4b umfassen die Förderung von Schlüsselqualifikationen im Unterricht des teilzeitschulischen Bereichs, der inneren und äußeren Differenzierung sowie der Untersuchung der Auswirkungen der Informationstechnologie auf Didaktik und Methodik des Unterrichts.

Stellt man nun die geförderten Modellversuche der Berufsbildungspolitik gegenüber und versteht man die in dem Bezugszeitraum zum Themenbereich 'Kooperation der Lernorte' geförderten Modellversuche als Möglichkeit, Reibungsverluste im dualen System zu minimieren, dann stellt dieser (Modellversuchs-) Ansatz den Versuch dar, die Einstellung der ausbildenden Betriebe gegenüber der Berufsausbildung

24) Zur Identifizierung der bildungspolitischen Intentionen siehe im Anhang die Anlagen zu Teil A der Arbeit.

insgesamt positiv zu beeinflussen und somit einen Beitrag zur Stärkung des Ausbildungsplatzangebotes zu leisten. Dieser Argumentationsstrang ist dann zutreffend, wenn man davon ausgeht, dass praktizierende oder potenzielle Auszubildende ihr Angebotsverhalten von der Qualität und Praxisnähe der Berufsausbildung leiten lassen. Sind finanzielle oder administrative Aspekte entscheidend für das Angebotsverhalten, so können Auszubildende auf diesem Wege nicht erreicht werden.

Durch Modellversuche zur inneren und äußeren Differenzierung werden einerseits Möglichkeiten evaluiert, Jugendliche aus den ‚Problemgruppen‘ in ihrer beruflichen Entwicklung zu fördern; in diesen Bereich sind vor allem auch die Modellversuche einzuordnen, die sich explizit mit der Entwicklung eines pädagogisch-didaktischen Konzepts für Jugendliche mit Lernschwächen beschäftigen (Förd.-Nr. 12/84-K 0641; 13/84, 9/85-K 0642, K 0642B). Andererseits wird die Differenzierung und die Begabtenförderung als Themenbereich aufgegriffen. Hierfür steht der Modellversuch, der sich mit der Entwicklung und Erprobung besonderer Maßnahmen für Jugendliche mit Hochschulzugang auseinandersetzt (Förd.-Nr. 41/87-K 0654).

Die Modellversuche zur Gruppe G4a - Innovationen in bestehenden Fächern oder Schulformen mit dem Schwerpunkt der Orientierung an den neuen Informations- und Kommunikationstechniken stehen in unmittelbarem Zusammenhang zum berufsbildungspolitischen Schwerpunkt 'Neue Technologien'. Hier wird demnach dieser Zielbereich, der zusätzlich noch durch die Änderung von Ausbildungsordnungen im allgemeinen berufsbildungspolitischen Bereich forciert wird, durch die Entwicklung didaktisch-methodischer Konzepte für den Schulunterricht (Gruppe G4b) oder durch die Untersuchung der Auswirkung dieser Technologien auf Didaktik und Methodik insgesamt weiter unterstützt.

Auch wenn keine Aussagen zum Ausmaß der Unterstützung der Berufsbildungspolitik durch die Modellversuchsarbeit möglich sind, so ist doch festzustellen, dass eine deutliche Kohärenz zwischen diesen Bereichen besteht. Die Aufgabe, die Modellversuchen in diesen Amtszeiten für die Berufsbildungspolitik zukommt, enthält deutlich unterstützende und innovative Momente.

Im zweiten Teil des Untersuchungszeitraumes beinhaltet die Gruppe G 4b folgende Aspekte: Die Ausbildung von Schlüsselqualifikationen und des selbstorganisierten Lernens mit einem Kernbereich in der Teilzeitberufsschule, einschließlich der Entwicklung ganzheitlicher Prüfungsaufgaben sind wesentliche Schwerpunkte, neben den Bereichen Europa, Differenzierung, Kundenorientierung und Umwelt, denen jeweils nur eine Nennung zuzuordnen ist.

Eine bedeutende Intention der Bundesminister in der Zeit 1992 bis 1998 war die Steigerung der Attraktivität des dualen Systems. Zwar ist die damit verbundene Zielrichtung, nämlich die Forcierung des Interesses von Schulabgängern an einer Berufsausbildung, vor allem auch derer mit Hochschulzugangsberechtigung, eindeutig; vielfältig dagegen sind die Maßnahmen, mit denen diese Intention verfolgt werden könnte. Dementsprechend weitläufig sind die Interpretationsmöglichkeiten, was die Kohärenz von Modellversuchsarbeit und Berufsbildungspolitik angeht.

Sieht man in der Ausbildung und Förderung von Schlüsselqualifikationen vor allem auch eine Anforderung an die berufliche Handlungskompetenz, wie sie zur Zeit von Seiten der Wirtschaft gefordert wird, dann sind Modellversuche zu diesem Bereich Möglichkeiten zur Kompetenzsteigerung des Lernortes Schule, was die Bedeutung dieses Partners für das duale System steigert und mithin auch die Attraktivität wachsen lässt. Dies setzt jedoch bei den Schulabgängern voraus, dass das fehlende Interesse am dualen System auf einen vermuteten Mangel an der Vermittlung eben dieses Kompetenzbereiches zurückzuführen ist. Wie das BMBF jedoch ausführt (s. o.), sind es in der Hauptsache die geringen Aufstiegs- oder Weiterbildungsmöglichkeiten aus der Berufsausbildung heraus, die die Hochschulzugangsberechtigten bewegen, ein Studium aufzunehmen bzw. die Berufsausbildung nur als Sprungbrett zu nutzen.

Diese Art der Interpretation ist von einer Vielzahl von Entscheidungsvariablen abhängig, so dass der Zusammenhang von Modellversuchsarbeit und Berufsbildungspolitik nicht eindeutig verifiziert werden kann. Gleiches gilt für die anderen Modellversuche aus der Gruppe 4b. Lediglich zwei weitere Erprobungsvorhaben, welche sich zum einen mit der Integration der europäischen Dimension in die Ausbildung der Europa-Wirtschaftsassistenten (Förd.-Nr.: 45/95, 46/95-K 2181, K 2181 B) und zum anderen mit der Differenzierung und Individualisierung befassen (Förd.-Nr.: 11/88-K 0240), fördern weitere Intentionen der Berufsbildungspolitik im Betrachtungszeitraum. Eine eindeutige Kohärenzaussage ist daher nicht möglich. Daraus lässt sich die These entwickeln, dass, wenn die betroffenen Bundesminister Modellversuche als unterstützendes Element ihrer Berufsbildungspolitik verstanden hätten, eine Zuordnung leichter herzustellen wäre. Der stützende Argumentationsstrang kann jedoch nicht entwickelt werden, da die vorgenommenen Zielformulierungen hinsichtlich der Struktur der politischen Intentionen nicht, bzw. zu wenig konkret entwickelt wurden. Auf Grund fehlender oder lückenhafter konzeptioneller Planungen, so die grundlegende Annahme (siehe hierzu auch die Ausführungen von Rolff 1971 u. Kap. 3.3.1), konnten die Modellversuche auch nicht zielgerichtet initiiert werden.

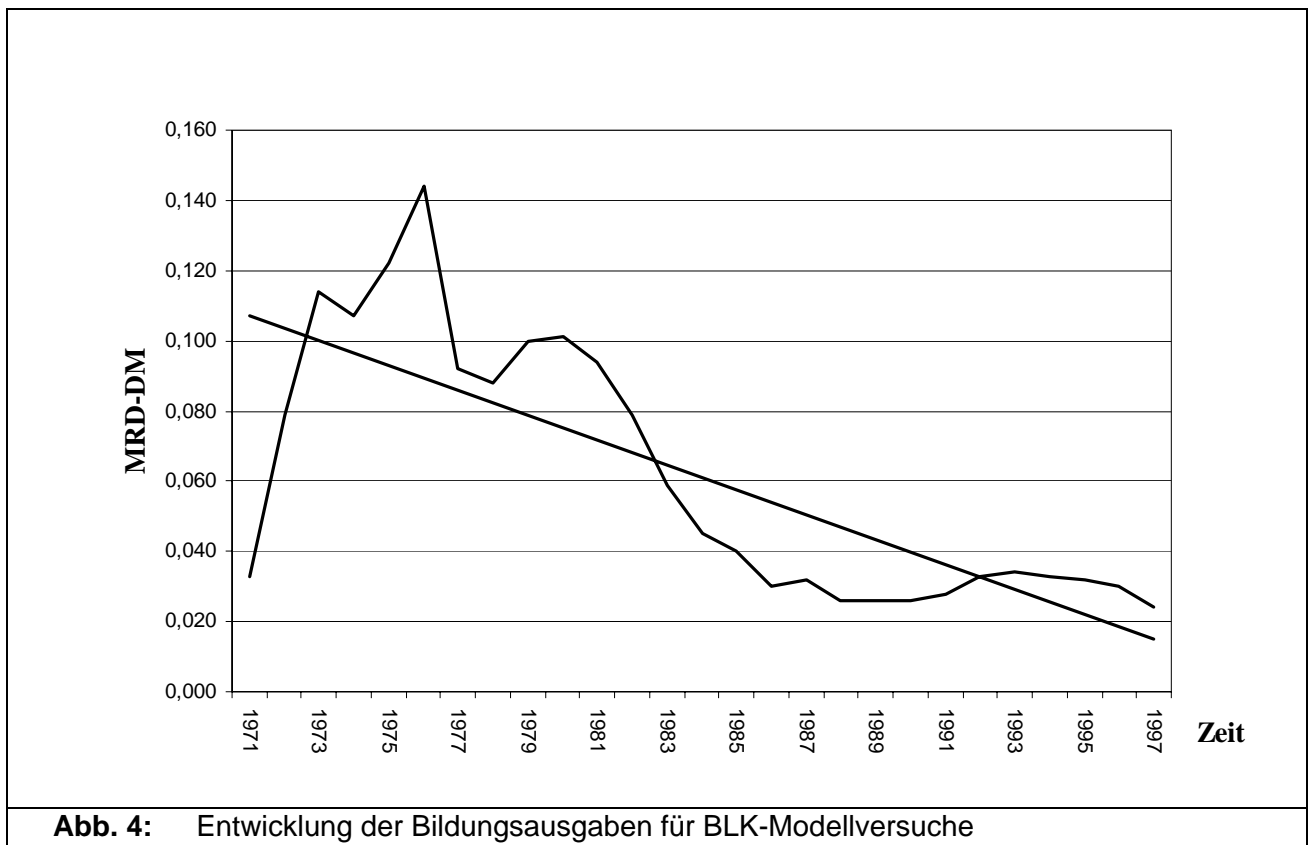
3.4 Chronologische Untersuchung der finanziellen Förderungsentwicklung

Die in Kap. 3.4.1 eingestellte Übersicht (Tab. 3) befasst sich mit der quantitativen Entwicklung der Fördermittel seitens des Bundes an BLK-Modellversuchen seit dem Gründungsjahr der Bund-Länder-Kommission 1971. Die Daten sind einerseits in absoluten Größen angegeben, andererseits ist auch deren relative Größe im Verhältnis zu anderen Bezugsgrößen ausgedrückt, um durch diese Vergleichsmaßstäbe Aussagen über die Entwicklung der Fördergrößen tätigen zu können. Bei der Betrachtung statistischer Erhebungsgrößen ist es sinnvoll, von einer hohen Betrachtungsebene schrittweise in eine tiefere zu wechseln, um so die Aussagefähigkeit einer Größe, hier der Bundesausgaben für BLK-Modellversuche, zu einer anderen Größe, Bruttoinlandsprodukt, Staatsverbrauch, Bildungsausgaben des Bundes, Ausgaben für Wissenschaft und Forschung außerhalb der Hochschulen, schrittweise zu entwickeln resp. zu steigern.

Des Weiteren soll auf die Problematik der Interpretation bzw. Aussagekraft solcher Datenzusammenstellungen eingegangen werden. Anhand von zwei Beispielen wird auf die Nicht-Statthaftigkeit der häufig vorgenommenen Darstellung von Zusammenhängen der Input- (DM) und Outputfaktoren (Leistung) eingegangen.

3.4.1 Darstellung der finanziellen Förderungsentwicklung von BLK-Modellversuchen²⁵

Im Zeitraum 1971 bis 1997 schwankt der Anteil der Bundesausgaben für das BLK-MV-Programm in erheblichem Maße (Tab.1/Spalte 1). Betrachtet man die Punkte der höchsten und niedrigsten Fördermittelausgaben so ergibt sich, dass im Tiefpunkt 1997 (24 Mio.) lediglich ein Achtel der 1976 (Höchstwert) ausgegebenen 144 Millionen Bundesmittel für Modellversuche verwendet wurden. Die Trendexploration verdeutlicht den rückläufigen Ausgabenverlauf:



Die Entwicklung des Bruttoinlandsproduktes (Tab.1/Spalte 2) verlief, abgesehen von zwei Einbrüchen 1975 und 1982, kontinuierlich; d. h. dieser Wohlstandsmesser konnte um ca. 88 % gesteigert werden.

25) Beabsichtigt sind in diesem Teil der Arbeit lediglich Trendaussagen zur Fördermittelentwicklung, so dass eine Abdiskontierung auf das Preisniveau eines Basisjahres unterbleiben kann. Dadurch entstehende Ungenauigkeiten beeinträchtigen die Aussagefähigkeit nicht.

Spalte	1	2	3	4	5	6	7	8	9	10	11
Jahr	Bundesmittel für BLK-Modellversuche in MRD DM	Bruttoinlandsprodukt in MRD DM	Verhältnis der Bundesmittel für MV am BIP in %	Veränderungen der MV-Ausg. am BIP ggü. Vorjahr in %	Staatsverbrauch in MRD DM	Nettoausgaben des Bundes für Bildung und Wissenschaft in MRD DM	Anteil der Bildungsausgaben am Staatsverbrauch in %	Verhältnis der Bundesmittel für MV an den Bildungsausgaben des Bundes in %	Veränderungen der MV-Ausg. an den Bildungsausgaben des Bundes in %	Ausg. f. Wiss. u. Forsch. außerhalb der Hochschulen in MRD DM	Verhältnis der Bundesausgaben für MV an den Ausg. für Wiss. u. Forsch. außerhalb der Hochschulen in %
1971	0,033	-	-	-	106,470	22,574	21,202	-	-	2,480	1,331
1972	0,079	1.658,0	0,0048	-	111,172	25,444	22,887	0,310	-	2,947	2,681
1973	0,114	1.737,0	0,0066	37,50	121,178	28,064	23,159	0,406	30,968	3,343	3,410
1974	0,107	1.740,0	0,0061	-7,58	134,909	49,351	36,581	0,217	-46,552	3,838	2,788
1975	0,122	1.718,6	0,0071	16,39	160,032	53,830	33,637	0,227	4,608	4,387	2,781
1976	0,144	1.810,1	0,0080	12,68	166,447	54,921	32,996	0,262	15,419	4,604	3,128
1977	0,092	1.861,6	0,0049	-38,75	174,104	57,411	32,975	0,160	-38,931	5,320	1,729
1978	0,088	1.917,4	0,0046	-6,12	191,221	61,301	32,058	0,144	-10,000	5,961	1,476
1979	0,100	1.998,4	0,0050	8,70	205,063	66,256	32,310	0,151	4,861	7,184	1,392
1980	0,101	2.018,0	0,0050	0,00	297,790	72,971	24,504	0,138	-8,609	7,666	1,318
1981	0,094	2.020,0	0,0047	-6,00	234,865	76,454	32,552	0,123	-10,870	8,071	1,165
1982	0,079	1.588,1	0,0050	6,38	246,617	77,486	31,420	0,102	-17,073	9,152	0,863
1983	0,059	2.036,2	0,0029	-42,00	336,210	78,061	23,218	0,076	-25,490	8,846	0,667
1984	0,045	2.093,5	0,0021	-27,59	350,230	77,572	22,149	0,058	-23,684	9,076	0,496
1985	0,040	2.136,0	0,0019	-9,52	365,720	80,765	22,084	0,050	-13,793	9,544	0,419
1986	0,030	2.186,1	0,0014	-26,32	382,550	80,523	21,049	0,037	-26,000	9,663	0,310
1987	0,032	2.218,4	0,0014	0,00	397,280	86,520	21,778	0,037	0,000	9,197	0,348
1988	0,026	2.301,0	0,0011	-21,43	412,380	88,181	21,383	0,029	-21,622	9,184	0,283
1989	0,026	2.384,4	0,0011	0,00	418,820	91,277	21,794	0,028	-3,448	9,173	0,283
1990	0,026	2.520,4	0,0010	-9,09	444,070	97,082	21,862	0,027	-3,571	9,634	0,270
1991	0,028	2.853,6	0,0010	0,00	556,720	107,823	19,368	0,026	-3,704	11,284	0,248
1992	0,033	2.916,4	0,0011	10,00	616,350	143,904	23,348	0,023	-11,538	12,109	0,273
1993	0,034	2.881,9	0,0012	9,09	634,860	154,607	24,353	0,022	-4,348	12,414	0,274
1994	0,033	2.960,2	0,0011	-8,33	658,580	157,695	23,945	0,021	-4,545	11,794	0,280
1995	0,032	3.013,8	0,0011	0,00	686,550	164,606	23,976	0,019	-9,524	11,770	0,272
1996	0,030	3.054,5	0,0010	-9,09	-	-	-	-	-	-	-
1997	0,024	3.121,2	0,0008	-20,00	-	-	-	-	-	-	-

Tab. 3: Entwicklung der Fördermittel des Bundes für BLK-Modellversuche

Bei der Betrachtung der Relation Entwicklung der Bundes-Modellversuchsausgaben und BIP ist ein rückläufiger Trend der MV-Ausgaben ab 1983 festzustellen (Tab.3/Spalte 3). Die prozentualen Veränderungen, (Tab.3/Spalte 4) ggü. dem jeweiligen Vorjahr, dokumentieren den gegensätzlichen Verlauf der BLK-Fördermittel zum BIP durch ihre negativen Steigerungsraten bzw. durch die Stagnation der Fördermittelentwicklung.

Als weitere Anhaltspunkte für die Ermittlung der Bedeutung des Instruments Modellversuch im Zeitablauf werden der Staatsverbrauch (Tab. 3/Spalte 5), die Bundesausgaben für Bildung und Wissenschaft (Tab. 3/Spalte 7) sowie die Ausgaben für Wissenschaft und Forschung außerhalb der Hochschulen (Tab. 3/Spalte 10) herangezogen, die den staatsbezogenen Teil der Verwendungsrechnung innerhalb der volkswirtschaftlichen Gesamtrechnung (VGR) wiedergeben, bzw. den Teil der Staatsausgaben dokumentieren, die auch die Modellversuchsmittel enthalten. Sie bieten somit die geforderten Bezugspunkte, damit die Relationen der Bildungsausgaben und der Modellversuchsmittel des Bundes ermittelt werden können.

Betrachtet man zunächst den Anteil der Bildungsausgaben am Staatsverbrauch (Tab. 3/Spalte 7), dann wird deutlich, dass dieser bis 1979 kontinuierlich zugenommen hat und durchschnittlich 30 % erreicht. Im weiteren Verlauf (ab 1981) nehmen die Bildungsausgaben ab, erreichen 1991 ihren Tiefpunkt. Beim letzten Bezugszeitpunkt liegt der Anteil bei ca. 24 % und stellt damit dennoch den größten Posten der Ausgabenrechnung (VGR) dar. In absoluten Zahlen haben sich die Nettoaussgaben des Bundes für Bildung und Wissenschaft erheblich vergrößert, was vordergründig für eine Ausweitung der Staatsinteressen an diesem Ausgabenbereich spricht (Tab. 3/Spalte 6).

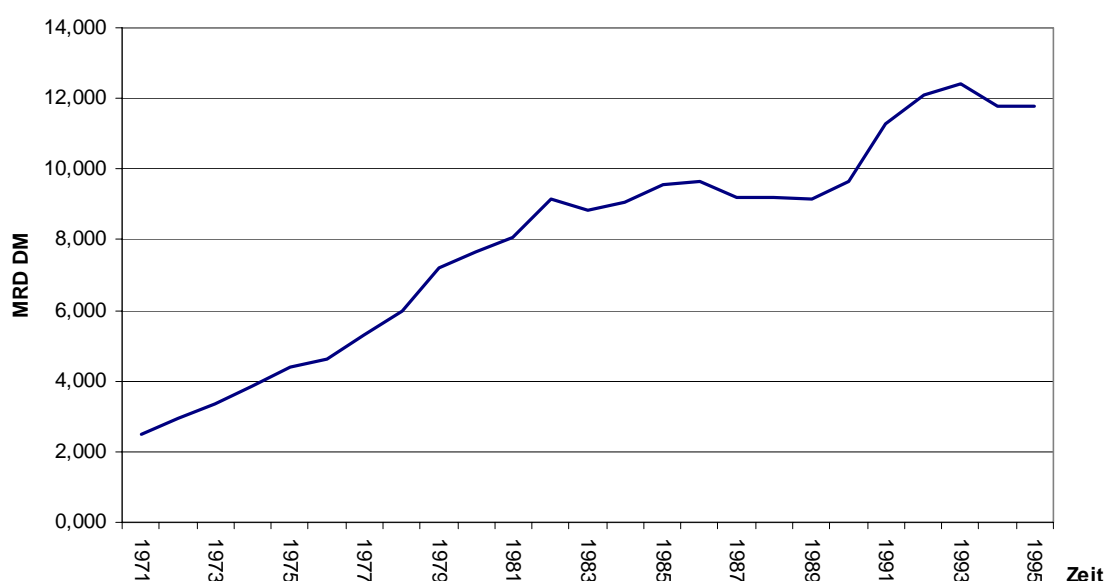


Abb. 5: Ausgaben für Wissenschaft und Forschung außerhalb der Hochschulen

Berücksichtigt man weiterhin den Anteil der Bundesmittel für MV an den Bildungsausgaben in Prozent (Tab. 3/Spalte 8), dann wird hier eine stark rückläufige Tendenz ersichtlich, die in einem schwachen Ausprägungsgrad (0,019 %) mündet. 1995 werden letztendlich noch 6,1 % vom ursprünglichen (1972=0,310 %) Anteil der MV-Mittel an den Bildungsausgaben des Bundes für dieses Instrument verwendet. Die Veränderungen ggü. dem jeweiligen Vorjahr (Tab. 3/Spalte 9) lassen negative Steigerungsraten mit wenigen Korrekturposten erkennen. Die Gründe für die genannten Korrektüreinschnitte ließen sich nur im gesamtpolitischen Zusammenhang erklären.

Insgesamt scheint sich eine Prioritätenverlagerung zugunsten anderer Aspekte eingestellt zu haben. Will man weitergehende Aussagen im Hinblick auf die Bedeutung der Modellversuche im chronologischen Verlauf treffen, so ist der Staatsverbrauch in einzelne Bestandteile weiter aufzuschlüsseln. So sind u. a. die Bundesausgaben für Modellversuche in den Ausgaben für Wissenschaft und Forschung außerhalb der Hochschulen enthalten (Tab. 3/Spalte 10). Diese Bezugsgröße hat sich im Betrachtungszeitraum um über 470 % gesteigert, während auch hier der Anteil für Modellversuche (Spalte 11) so kontinuierlich abgenommen hat, dass er 1995 lediglich noch 0,27 % einnahm; selbst im Jahr der Initiierung der Rahmenvereinbarung Modellversuche sind es nur 1,3 % gewesen.

In diesem Zusammenhang ist es fraglich, ob ein Instrument wie das der Modellversuche, das mit so einem geringen finanziellen Förderaufwand ausgestattet ist, noch als Mittel anzusehen ist, das Reformen und wesentliche Impulse im Bildungswesen hervorbringen kann und ob dies politisch überhaupt gewollt ist. Aufgrund der dargestellten Finanzsituation sind hieran erhebliche Zweifel anzumelden.

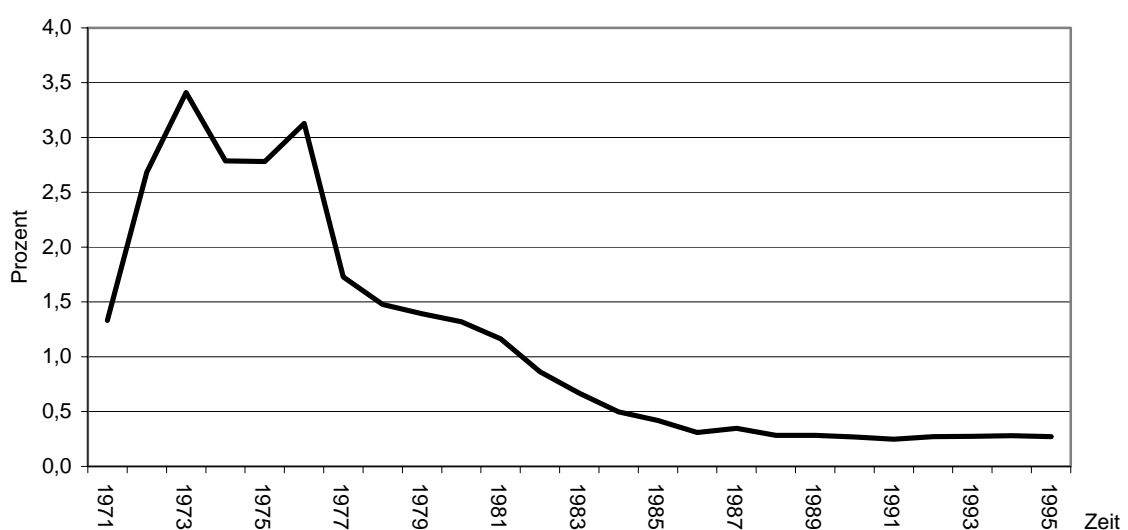


Abb. 6: Anteil der Bundes-MV-Ausgaben an den Ausgaben für Wissenschaft und Forschung außerhalb der Hochschulen

3.4.2 Darstellungs- und Interpretationsprobleme im Hinblick auf den Input-, Outputzusammenhang

Die vorangegangenen Ausführungen nähren die Vermutung, dass die Bedeutung des Instruments Modellversuch im Zeitablauf innerhalb der (bildungs-) politischen Prioritätenliste abgenommen hat. Auch in der Alltagsdiskussion und in der Literatur wird häufig ein direkter Kausalzusammenhang zwischen Inputfaktoren eines Instruments und der Leistungserbringung für den beabsichtigten Wirkungsbereich unterstellt. In dem hier vorliegenden Diskussionszusammenhang würde dies bedeuten, dass die Transferrate der Modellversuche als übergeordnete Zielgröße, bzw. die spezifischen evaluations- oder modellversuchsbezogenen Ziele durch die Erhöhung der monetären Mittel gesteigert werden könnten. Demnach würden der Transfer (oder andere EV/MV-Ziele) als abhängige Variable und als Outputfaktor zur finanziellen Förderung definiert; die Ubiquität dieses Argumentationsstranges soll an zwei Beispielen aus der Literatur nachgewiesen werden.

Die Bund-Länder-Kommission stellt in ihrer Veröffentlichung '10 Jahre Modellversuche im Bildungswesen' fest: "In Anbetracht des finanziellen Aufwandes für Modellversuche ist es verständlich, dass häufig und bisweilen auch mit einer gewissen Ungeduld nach ihren Ergebnissen gefragt wird. Hier hat die Erfahrung gezeigt, dass Modellversuche in aller Regel weit weniger spektakuläre Ergebnisse erbringen, als dies ursprünglich wohl von den Beteiligten und der Öffentlichkeit erwartet worden ist" (BLK 1981: 6).

Während im ersten Satz des Zitates nur eine Erwartungshaltung zum Ausdruck gebracht wird, stellt der Autor mit dem zweiten Satz eine direkte Verbindung von 'finanziellem Aufwand' und 'spektakulären Ergebnissen' her. Selbst wenn der Output hier nicht genauer definiert wird, ist die Verbindung von Einsatz- zu Ausbringungsfaktoren deutlich. Noch stärker wird der unterstellte Kausalzusammenhang in einem Aufsatz von Horst Weishaupt deutlich, der die finanziellen Ressourcen der Bildungsforschung im Hinblick auf die Ausgabenentwicklung außeruniversitärer Einrichtungen und der Entwicklung der Drittmittel für Grundlagenforschung untersucht hat.

Hier wird resümiert: "Dass Maßnahmen dringend sind, um die universitäre Erziehungswissenschaft nicht wieder gänzlich zum literarisch-spekulativen Bildungsfach werden zu lassen, ist aus dem Rückgang der Drittmittel vor dem Hintergrund eines traditionell niedrigen Forschungsniveaus beziehungsweise der häufig unzureichenden Professionalität der Forschung begründbar" (Weishaupt 1985: 99).

Anders formuliert erklärt Weishaupt, dass das Niveau der Erziehungswissenschaft insgesamt und das Forschungsniveau sowie die Professionalität der Forschung im Besonderen durch die Erhöhung der Drittmittel gesteigert werden könne. Auch hier wird somit ein direkter Zusammenhang von Inputfaktoren (Drittmittel) und Outputfaktoren (Qualität der Erziehungswissenschaft, Forschungsniveau und Professionalität) unterstellt.

Diese Form der Argumentation ist zwar naheliegend, impliziert jedoch automatisch, dass eine Erhöhung der Inputfaktoren zu einer Verbesserung derjenigen Einflussfaktoren führt, die auf das Outputniveau wirken, was dadurch erhöht wird. Dabei wird vernachlässigt, dass sich diese Wirkung zwar einstellen kann, dass eine Wirkungsverbesserung jedoch nicht zwingend ist, da die Inputfaktoren nicht automatisch oder in jedem Fall an die Stellen im betrachtenden System gelangen, die auch Einfluss auf eine qualitative oder quantitative Erhöhung der Outputfaktoren nehmen können. Darüber hinaus ist jedoch die Zahl der intervenierenden Variablen so groß, dass dieser Zusammenhang nicht als klar ausgewiesen dargestellt werden darf; diese Form der Argumentation ist nicht haltbar.

Demnach kann das oben abgebildete Zahlenmaterial (siehe Kap. 3.4.1) nicht derart gedeutet werden, dass die Abnahme der Fördermittel für Modellversuche automatisch zu einer sinkenden Leistungserbringung für die beabsichtigten Zielbereiche führt. So wäre es denkbar, dass die Handlungsfähigkeit von Schülern bei abnehmenden Modellversuchsmitteln gesteigert wird, wenn die verbleibenden finanziellen Ressourcen an die Stellen des Bildungsprozesses gelangen, die stark einflusswirksam sind, wie z. B. die Ausbildung des Lehrpersonals und auf andere Faktoren, wie eine hochmoderne und kostenintensive Schulausstattung verzichtet würde. Allerdings ist auch dieser Alternativaspekt nicht belegbar.

Im Ergebnis lässt sich festhalten, dass die Darstellung der Ausgabenentwicklung der Bundesmittel für Modellversuche zwar unter dokumentarischen Gesichtspunkten oder für politische Interessengruppen von Bedeutung sein kann, dass eine Interpretation solcher Größen jedoch nicht möglich ist im Hinblick auf Rückschlüsse der Leistungserbringung von Modellversuchen für Bildungsplanung, -politik oder -praxis. Es sind lediglich Indizien gegeben, die durch die Kombination mit anderen Erkenntnissen gestützt oder widerlegt werden können.

Hierzu dienen die Ausführungen zur Kohärenz von Modellversuchen und Berufsbildungspolitik sowie die nachfolgende metaevaluative Studie zu den Leistungen durchgeführter Modellversuche (siehe Kap. 6).

4 Zum Stand der Evaluationsforschung

4.1 Der Begriff 'Evaluation'

Durch die intensive Auseinandersetzung mit dem Themengebiet 'Evaluation' im angelsächsischen Raum, insbesondere durch die Entwicklung der Theorie von Tyler (congruence between performance and objectives), sind wesentliche Grundlagen und Entwicklungslinien für dieses Wissenschaftsgebiet generiert worden. Nach Ansicht von Grüner ist im deutschsprachigen Raum kein entsprechendes Pendant, keine eigene Forschungstradition hierzu festzustellen (vgl. Grüner 1995: 382); Gegenstand ist hier weniger eine evaluationsbezogene Fragestellung, sondern vielmehr die Auseinandersetzung mit dem Aspekt der Wissenschaft-Praxis-Kommunikation, die Nähe der wissenschaftliche Begleitung zum Praxisfeld (Modellversuchsfeld). Im Hinblick auf die Entwicklung eines Evaluationsrasters, das auf der Metaebene die Güte von Modellversuchsevaluationen (Konstruktionen zweiter Ordnung; siehe Kap. 5.1) einschätzen soll, ist es daher zweckdienlich, einschlägige evaluationsbezogene Literatur im Hinblick auf enthaltene Themenschwerpunkte auszuwerten und von der Literatur des deutschsprachigen Raums im Wesentlichen Abstand zu nehmen.

Die folgenden Definitionen geben einen Überblick über die Bandbreite des Evaluationsbegriffs, wie er in der angelsächsischen Literatur Verwendung findet; darüber hinaus bilden sie die Grundlage für eine 'Gegenstandsuntersuchung':

- "The process of evaluation is essentially the process of determining to what extent the educational objectives are actually being realized by the program of curriculum and instruction" (Tyler 1950: 69).
- "Evaluation is by definition a continuous process with a life cycle that parallels the educational product life cycle. It is a means for enhancing the use of available resources by providing the information required for managing an educational service function" (Arnoff 1987: 104).
- "Evaluation is the systematic collection of descriptive and judgemental information necessary to make effective decisions related to the selection, adoption, value, and modification of various training activities. [...] From this perspective, evaluation is an information gathering technique" (Goldstein 1987: 140).
- "Systematic educational evaluation consists of a formal appraisal of the quality of educational phenomena" (Popham 1988: 7).
- "[...] program evaluation is a collection of methods, skills, and sensitivities necessary to determine whether a human service is needed and likely to be used, whether it is sufficiently intense to meet the need identified, whether the service is offered as planned, and whether the human service actually does help people in need without undesirable side effects" (Posavac, Carey 1989: 3).

- "Evaluation as we know it did not simply appear one day; it is the result of a developmental process of construction and reconstruction that involves a number of interacting influences" (Guba, Lincoln: 1989: 22).
- "The systematic process of collection, analyzing, and interpreting information to determine the extent to which pupils are achieving instructional objectives (Answer the question 'How good?)" (Gronlund, Linn: 1990).
- "Evaluation is the systematic assessment of the worth or merit of some object. [...] According to this definition, if a study does not report how good or bad something is, it is not evaluation" (Stufflebeam, Shinkfield 1990: 3).
- "Collection and use of information to make decisions about an educational program" (Cronbach 1991: 101f).
- "[...] the process of determining the merit, worth and value of things, and evaluations are the products of that process" (Scriven 1991: 1).
- "Evaluation is the systematic investigation of the merit and worth of an object (program) for purpose of reducing uncertainty in decision making" (Mertens 2000: 41).
- "Professional evaluation is essentially about generation information that assists others in making judgements about a program, service, policy, organization, person, or whatever else is being evaluated" (Henry 2001: 1).

Grüner, der einige der oben genannten Definitionen auf Gemeinsamkeiten hin untersucht, kommt zu der Überzeugung, dass folgende fünf Kennzeichen in allen Definitionen von Evaluation wieder zu finden seien - Zielorientierung, Systematik, Kontinuität, Bewertung und Objektorientierung:

Eine Evaluation sei immer auf ein Ziel hin ausgerichtet, das in der Verbesserung des Untersuchungsobjekts liege; die Gewinnung der Daten habe systematisch zu erfolgen, wobei es von besonderer Bedeutung sei, dass das Konzept und die Systematik der Datengewinnung während des Evaluationsprozesses kontinuierlich beizubehalten ist. Die gewonnenen Daten müssten interpretiert und im Hinblick auf die Güte des Objekts bewertet werden; Gegenstand der Evaluation sei außerdem ein identifiziertes Objekt, d. h. eine Maßnahme oder ein Programm (vgl. Grüner 1995: 384).

Ergänzt werden könnte diese Merkmalsaufzählung um den Aspekt der Prozessorientierung, der in dem definitionsimmanenten Zusammenhang von Datensammlung, -analyse und -interpretation zum Ausdruck kommt; im Hinblick auf die Generierung aussagefähiger Evaluationsergebnisse ist dabei einerseits die Bedeutung der stringenten Trennung und andererseits die der gegenseitigen Abhängigkeit dieser Evaluationsschritte hervorzuheben.

Neben dieser Aufzählung von charakteristischen Merkmalen des Begriffes 'Evaluation' bestehen auch andere, wie z. B. die von Bortz und Döring (1995), Cronbach (1972), Rossi und Freeman (1995) sowie Scriven (1972). Merkmalskataloge weisen jedoch wegen ihrer nicht vorhandenen Endlichkeit keine hinreichende Eindeutigkeit auf, um den Gegenstand 'Evaluation' zu beschreiben. Daher soll nachfolgend die Konzentration auf einen 'Minimalkonsens der scientific community', wie er von Wottawa/Thierau aufgezeigt wird, erfolgen. Ziel ist es dabei nicht, eine omnipotente Definition zu generieren, sondern es soll aufgezeigt werden, wie der Minimalkonsens ausgestaltet bzw. um welche Inhalte (Bausteine/Module) dieser ergänzt werden muss, um für den spezifischen Evaluationsgegenstand (Wirkungsorientierung s. u.) Gültigkeit zu erlangen.

Wottawa/Thierau heben hervor, dass Konsens dahingehend bestehe, dass a) wissenschaftliche Evaluation etwas mit 'Bewerten' zu tun habe und als Planungs- und Entscheidungshilfe zur Bewertung von Handlungsalternativen diene. Darüber hinaus sei sie b) ziel- und zweckorientiert und es bestehe im wissenschaftlichen Sprachgebrauch Einigkeit darüber, dass für Evaluationsmaßnahmen c) wissenschaftliche 'Forschungsmethoden und -techniken' angewandt werden sollten, die dem 'aktuellen Stand' entsprechen (vgl. Wottawa, Thierau 1998: 14).

Für die weitere Begriffsentwicklung wird daher die Verwendung einer Basisdefinition/Ausgangsdefinition vorgeschlagen, die zunächst einen Minimalkonsens in Bezug auf den von Wottawa/Thierau genannten Aspekt a) 'Bewertung' darstellt. Hierfür eignet sich der erste Teil der Definition von Mertens (s. o.):

"Evaluation ist eine systematische Untersuchung der Leistungen und des Wertes eines Objektes (Programmes) [...]" (Mertens 2000: 41).

Des Weiteren muss deutlich werden, was genau 'Gegenstand der Wertzumessung' beim Zielobjekt ist (b - Ziel-, Zweckorientierung s. o.). In diesem Zusammenhang ist auf den von Scriven geprägten Ausdruck 'evaluand' hinzuweisen, der zum Ausdruck bringen soll, dass es eine Vielzahl von Evaluationsgegenständen innerhalb eines Evaluationsobjektes gibt (vgl. Scriven 1973: 324).

In Evaluationen können demnach verschiedene Bestandteile und Ergebnisse von Modellversuchen als 'Evaluand' identifiziert werden, deren Leistungen sich an unterschiedlichen Stellen der 'Wirkungskette', auf unterschiedlichen Wirkungsebenen zeigen: beim sog. 'output' (Leistungserbringung) kann es sich einerseits um 'interne' oder 'externe' Produkte handeln, d.h. zum Beispiel um das Modellversuchskonzept (interner Bezug) oder um das zu erstellende Curriculum (externer Bezug). Des Weiteren kann der bei Personen ermittelbare 'outcome' (Wirkungen bei den direkten Adressaten) und der in Folgeprogrammen, dann in kodifizierter Form (z. B. in Lehrplänen oder Planungsmodellversuche), oder bei nicht direkt in den Modellversuch einbezogenen Personen feststellbare 'impact' (gesellschaftliche Auswirkungen) Gegenstand von Evaluationen sein (siehe hierzu auch Widmer 2000: 78).

Die nachfolgende 'Evaluand-Wirkungs-Matrix' lenkt den Fokus bei der Begriffsbestimmung von Evaluation auf den Gegenstand der Wertzumessung, resp. auf die Stelle in der Wirkungskette (Wirkungsebene), an der die Leistung des Evaluand gemessen werden soll. Die 'Evaluand-Wirkungs-Matrix' soll dabei verdeutlichen, dass die verschiedenen Evaluations-Gegenstände von Modellversuchen Ergebnisse/Erkenntnisse (Leistungen) darstellen, die Auswirkungen auf interne oder externe Personen, Prozesse oder neue Modellversuche haben können, welche auf verschiedenen Wirkungsebenen zum Tragen kommen:

	Evaluand		Generierte Informationen	Wirkungs-entfaltung
Modellversuchsfeld (Explikationsmodellversuch)	Modellversuchskonstrukt/-design/-treatment		Aufbau und Inhalte des Modellversuchs	Output
	Maßnahmen der Modellversuchsimulation		Kongruenz von Modell-'Umwelt' und tatsächlicher 'Umwelt'	
	Modellversuchsprozesse		Informationen über kausale Prozesse zwischen MV-Treatment und MV-Wirkungen	
	Lernleistungen/ Kompetenzen der Probanden		Kompetenzausprägungen	Outcome
Funktionsfeld (Planungsmodellversuch; Bildungspraxis)	Transfer	Kodifizierungen	Auffinden von Modellversuchserkenntnissen in Dokumentationen – z. B. in RL, RRL o. ä.	Impact
		Transferprozess	Informationen über Generalisierung, z. B. Struktur, Arbeitsweisen von Explikationsorganisationen und Planungsorganisationen, Einflussfaktoren auf das Treatment	

Abb. 7: Evaluand-Wirkungs-Matrix

Die aufgezeigte Bandbreite von Evaluationsgegenständen (Evaluands) eines Projektes, wie z. B. eines Modellversuches, verdeutlicht, dass eine alle Wirkungsebenen umfassende Evaluation zumindest eine interne Sequenzierung und Ausgestaltung erfordert, wenn nicht sogar eine generelle Beschränkung auf eine Wirkungs-

ebene sinnvoll ist, da diese unterschiedliche Evaluationskonzepte, methodische Herangehensweisen u. a. notwendig macht.

Weiteres Merkmal von Evaluation und Bestandteil einer auf einem Minimalkonsens beruhenden Begriffsbeschreibung ist die Verwendung 'wissenschaftlicher Techniken' (Evaluationsmethoden), die dem 'aktuellen Stand' entsprechen (Merkmal c; s. o.). Widmer stellt generell für die Evaluationsmethodologie fest, dass sich diese (auch) durch ihre Vielfalt an methodischen Ansätzen auszeichne, die zum Einsatz kommen können. Heute seien diese Ansätze durch ihre weitgehend methodische und erkenntnistheoretische Offenheit charakterisiert. Dies gelte sowohl auf der Ebene der epistemologischen Grundannahmen, des Untersuchungsdesigns (vgl. Klöti; Widmer 1997: 26) als auch für die Methodik der Datenerhebung und Datenauswertung (vgl. Widmer; Binder 1997: 201; Widmer 2000: 80)²⁶.

Für eine auf ein bestimmtes Anwendungsfeld gerichtete Evaluationsdefinition, d. h. auf eine ausgewählte Wirkungsebene zielende Definition ist daher auch der methodische Aspekt zu spezifizieren und in die jeweils 'wirkungsbezogene' Definition aufzunehmen.

Behält man den Fokus der Wirkungsorientierung bei, dann ist m. E. der Minimalkonsens dahin gehend zu ergänzen, dass zusätzlich zu den Fragen, was evaluiert wird (b Evaluand – Ziel-, Zweckorientierung) und welche Evaluationsmethoden (c wiss. Forschungsmethoden, -techniken) zum Einsatz kommen, abschließend zur Begriffsbestimmung auch auf den Aspekt des Zeit-Raum-Verhältnisses einzugehen, der deutlich macht, zu welchem Zeitpunkt und in welchem 'Verwendungsraum/-zusammenhang' die Evaluationsergebnisse genutzt werden können. Widmer hat hierfür den Begriff der 'Verortung einer Evaluation' geschaffen (vgl. Widmer 2000: 79). Während eine formative Evaluation zeitlich gesehen parallel zum Modellversuchsprozess durchgeführt wird und Impulse/Korrekturen für den weiteren Verlauf gibt (Modellversuchsraum), wird die summative Evaluation ex-post initiiert; ihre Wirkung ist somit prospektiv und kann somit Einfluss auf den nachfolgenden 'Modellversuchsraum' (Planungsmodellversuch) bzw. auf die Bildungspraxis nehmen.

Zusammenfassend kann der modulare Evaluationsbegriff folgendermaßen abgebildet werden:

Basisdefinition +				
Baustein A: Gegenstands-/ Wirkungsdimension	+	Baustein B: methodische Dimension	+	Baustein C: Raum-Zeit-Dimension

Die obige Darstellung betont den modularen Charakter von Evaluation; durch Spezifikation im Hinblick auf Gegenstand und die methodische Herangehensweise so-

26) Zur Entwicklung der Evaluationstheorie und –methodik siehe Kap. 4.2-4.4.

wie durch die Angabe der Raum-Zeit-Dimension ergibt sich eine auf den Wirkungsbereich (Bewertungsaspekt) bezogene Begriffsbeschreibung, die verdeutlicht, dass die Evaluations-Module nicht unabhängig voneinander zu betrachten sind, sondern dass diese eine Kongruenz aufweisen: die methodischen Dimension und die Raum-Zeit-Dimension sind im Hinblick auf ihre Ausgestaltung vom Evaluationsgegenstand (Gegenstandsdimension) abhängig, so dass vom Primat des Evaluands in einer Evaluationsdefinition gesprochen werden kann.

Zur weiteren Begriffsverdeutlichung und bevor in Kapitel 5 eine Beschreibung des eigenen Vorgehens und damit eine für diese Arbeit gültige Evaluationsdefinition erfolgen kann, ist eine Abgrenzung zum Forschungsbegriff notwendig, da die Begriffe Evaluation und Evaluationsforschung häufig parallel resp. synonym verwandt werden, wie z. B. von Rossi und Freeman:

"Evaluation research is the systematic application of social research procedures in assessing the conceptualization and design, implementation, and utility of social intervening programs" (Rossi, Freeman 1986: 19).

Nach Ansicht von Beywl jedoch unterscheiden sich Evaluation und (Evaluations-) Forschung in ihren Zielsetzungen. Während Evaluation auf die Beschaffung praxisrelevanter Informationen gerichtet sei, zielt Forschung auf die Vermehrung gesicherten Wissens ab (vgl. Beywl 1991: 267). Außerdem sei Evaluation durch die Informationssammlung und die darauf aufbauende Bewertung von Handlungsalternativen entscheidungsorientiert angelegt und untersuche Problembereiche, um zur Entscheidungsfindung beizutragen, während (Bildungs-) Forschung auf Konklusionen ausgerichtet sei (vgl. Grüner 1993: 49).

Neben dem Zielaspekt würden sich Evaluation und (Bildungs-) Forschung außerdem durch das Bestreben nach Generalisierbarkeit und durch den Bewertungsaspekt unterscheiden. Eine Evaluation untersuche ein singuläres Phänomen, eine Bildungsmaßnahme, wie z. B. einen Modellversuch und verfolge damit nicht die Absicht, eine Verbindung zu anderen Erkenntnissen herzustellen und hierdurch zu generalisieren. Darüber hinaus würden Evaluationen Daten interpretieren und bewerten, während Bildungsforschung auf die Generierung möglichst objektiver Daten ausgerichtet sei und diese wertfrei behandle (vgl. Grüner 1995: 396f).

Popham fasst die von ihm als relevant erachteten Unterscheidungsmerkmale in einer Übersicht zusammen, die hier durch den von Beywl hervorgehobenen Aspekt der 'Zielsetzungen' ergänzt wird (vgl. Popham 1988: 13):

Untersuchungsmerkmale	Evaluation von Bildungsmaßnahmen	Bildungsforschung
Zielsetzung	Beschaffung praxisrelevanter Informationen	Vermehrung gesicherten Wissens ²⁷
Fokus	Entscheidung	Folgerungen (conclusions)
Generalisierbarkeit	Niedrig	hoch
Ausrichtung (value emphasis)	Wert (worth)	Wahrheit (truth)

Abb. 8: Abgrenzung von Evaluation und Bildungsforschung

Die o. g. Unterscheidungsmerkmale unterstreichen den Fokus der Wirkungsorientierung in der Definition des Begriffes 'Evaluation', da sie die Betonung des Nutzens von Evaluation als 'Maßnahme zur Entscheidungsfindung' hervorheben und damit verdeutlichen, dass ausgewählte Aspekte der Wirkungskette hinsichtlich ihrer Leistung eingeschätzt werden, ohne das Ziel zu verfolgen, allgemeingültige Erkenntnisse zu generieren. Die nachfolgende Darstellung dokumentiert einerseits den Stand der Auseinandersetzung der Evaluationsforschung unter dem Aspekt der Methodologiediskussion (chronologische und typologische Darstellung) und der Einschätzung theoriegeleiteter Evaluationen als Kernbereiche der wissenschaftlichen Diskussion in diesem Themengebiet, so wie er in der Literatur wiedergegeben wird, andererseits trägt sie zur Spezifizierung des methodischen Bausteins (methodische Dimension) in der für diese Arbeit geltenden Evaluationsdefinition bei.

4.2 Entwicklungslinien in der Evaluationsmethodologie der nordamerikanischen Forschungstradition

4.2.1 Chronologische Entwicklung der Evaluationsmethodologie

Methodologie kann zunächst als "[...] zusammenfassende Lehre oder Theorie von den Methoden, die zur Erreichung bestimmter Ziele in Erziehung und Unterricht zur Verfügung stehen" bezeichnet werden; sie "[...] befasst sich mit der empirischen Überprüfung, theoretischen Begründung und Weiterentwicklung von Methoden" (Schaub 1995: 246).

Beywl lehnt sich bei seiner Beschreibung des Begriffes 'Methodologie' u. a. an Opp an und hält fest, dass methodologische Aussagen Kriterien und Regeln seien, mit denen sozialwissenschaftliche Praxis kritisch überprüft werden könne und darüber hinaus diese Kriterien einerseits als Orientierungshilfe für die wissenschaftliche Arbeit angesehen werden müssten, andererseits die wissenschaftliche Tätigkeiten auf die Realisierung dieser Kriterien ausgerichtet sein sollte.

Aus dieser Beschreibung resultiert eine Doppelfunktion von Methodologie, die erstens in der Ermöglichung der Bewertung wissenschaftlichen Handelns (kritisch-

27) Das 'gesicherte Wissen' muss im Hinblick auf den Zeitfaktor und das Problem eingeschränkter Transparenz über den 'Stand des Wissens' mit dem Attribut der 'vorläufigen Gültigkeit' versehen werden. Gleiches gilt für 'Wahrheit' als Implikation der Bildungsforschung; in diesem Zusammenhang erscheint es angebracht, den von 'Popham' verwendeten Begriff der 'Wahrheit' durch '(vorläufige) Gültigkeit' zu ersetzen.

analytische Funktion) und zweitens in der Anleitung desselben (konstruktiv-normative Funktion) zu sehen ist (vgl. Beywl 1988: 10; siehe auch Opp 1970: 15). In der Literatur sind verschiedene Phasenverläufe zu finden, die die Entwicklungslinien in der Evaluationsmethodologie nachzuzeichnen versuchen. Dabei orientiert man sich u. a. an einer chronologischen Phaseneinteilung, denen die verschiedenen Evaluationsansätze zugeordnet werden, wobei die Abgrenzung der Zeitabschnitte nicht immer trennscharf erfolgen kann und somit die Einstufung von Evaluationsansätzen problematisch ist (siehe hierzu Stufflebeam, Shinkfield 1990: 16ff; Guba, Lincoln 1989: 22ff und hierauf aufbauend Mertens 2000: 43ff).

Die nachfolgende Darstellung dient daher auch nur der chronologischen Orientierung, bevor in Kap. 4.2.2 zu den Inhaltsaspekten Bezug genommen wird; die Übersicht zeigt eine Zusammenfassung der von Beywl dokumentierten Entwicklung der Evaluationsmethodologie, in die zeitgeschichtliche und soziale Bezüge eingeflossen sind:

Phase	Zeitraum	Ausprägung
1	bis 1925	Vorläufer – Akkreditation und Testverfahren
2	1925 – 1945	Grundlegung – Curriculumevaluation
3	1945 – 1957	Stagnation – Rückgriff auf qualifizierte Verfahren
4	1957 – 1970	Boom – Expansion der Evaluationsbranche und methodologische Innovation
5	1970 – 1980	Verunsicherung – Nutzendefizit und Werte-Daten-Dilemma
6	ab 1980	Herausforderung – Bewährungsprobe der Evaluation im Schatten der Haushaltskonsolidierung

Abb. 9: Phasen der Evaluationsmethodologie nach Beywl

4.2.2 Typologische Darstellung der Evaluationsansätze

Wegen der Vielzahl der Evaluationsansätze, die im Zeitablauf entstanden sind, bestehen auf methodologischem Gebiet seit den 70-er Jahren Versuche, diese Ansätze systematisch zu vergleichen, um so einen Beitrag zur Weiterentwicklung der Evaluationsforschung und –praxis zu leisten. Trotz des Problems der Zuordnung von Ansätzen zu Modelltypen aufgrund von 'Unschärfen' bei der Evaluationskonzeption und der regelmäßigen Weiterentwicklung von Evaluationsansätzen ist eine Typologie sinnvoll, um die Vergleichbarkeit der Modelltypen herzustellen und die Anwendungsbereiche von Modelltypen aufzuzeigen bzw. die Diskussion hierzu zu ermöglichen resp. zu erleichtern. Die im Folgenden dargestellte und von Beywl entlehnte Typologie von Evaluationsmodellen baut auf den Arbeiten von House (1980) und Stufflebeam, Webster (1981) auf; durch die Weiterentwicklung schon bestehender Typologien und durch die im Zeitablauf vorgenommene Auseinandersetzung der Literatur mit diesen Erkenntnissen, kann von einer 'kommunikativen Validierung' ausgegangen und Beywls Typologie als ausreichend gefestigt angesehen werden. Die Charakterisierung der verschiedenen 'Evaluationstypen' erfolgt nach dem Muster Ziel-, Gegenstandsbeschreibung, methodische Implikationen, Bezugnahme auf einen für den jeweiligen Bereich beispielhaften

nahme auf einen für den jeweiligen Bereich beispielhaften Evaluationsansatz, bedeutende Vertreter sowie Chancen und Risiken:

Effizienzkriterien-gesteuerte Evaluationsmodelle,

nehmen einen quantitativen Input-Output-Vergleich vor, um die Effizienz und Effektivität eines Programms oder Curriculums zu messen (vgl. Beywl 1988: 45).

Zu Beginn der Evaluation werden einzelne Ergebnisvariablen definiert und operationalisiert sowie gemessen und in Beziehung zu Programmalternativen gesetzt, um so den 'Opportunitätsgewinn/ -verlust' ermitteln zu können. Charakteristisch für den Evaluationsprozess dieser Modelle im sozialwissenschaftlichen Tätigkeitsbereich sind Überlegungen zur Optimierung der Effizienz. Ziel der Evaluation ist es, diejenigen Einflussvariablen offen zu legen, die z. B. die Lernleistung in Bezug auf die gewünschten Ergebnisse (Outcomes) maximieren (vgl. Thorndyke 1971). Die Erreichung der Programmziele (Output) wird ebenso ermittelt wie in den Modellen der Programm-, Lernzielgesteuerten Evaluation.

Entgegen der wirtschaftswissenschaftlichen Variante, die mit monetären Größen arbeiten und vollständig in den ökonomischen Zweckhorizont eingebunden ist, stellen die sozialwissenschaftlichen Evaluationsansätze Verhaltensvariablen in den Vordergrund "[...] deren Beeinflussung durch Programmintervention überprüft werden soll, wobei die Effizienz der Einwirkung das herausragende Kriterium ist" (Beywl 1988: 49).

Rossi und Freeman, die den diesem Modelltypus zugehörigen 'impact-' Ansatz entwickelt haben, gehen davon aus, dass ein Programm (z. B. Modellversuch) das Potenzial besitzt, Veränderungen in Menschen hervorzurufen (\approx impact), die von dem Programm beeinflusst werden (vgl. Houston 1972: 51).

Darüber hinaus kann der Ansatz als ein Versuch beschrieben werden, Annahmen über Regulierung, Modifikation und Kontrolle sozialen Verhaltens in Hypothesen zu übertragen (Kausalhypothese, Interventionshypothese, Aktionshypothese; vgl. Rossi, Freeman 1982: 62f), auf die rationales Handeln bezogen werden kann. Die Entwicklung der Hypothesen erfolgt aus 'experimentellen Studien oder gut entwickelten Theorien'. Optimal sei es, wenn das Interventionsprogramm wissenschaftlich geplant würde, um so die Hypothesen des Programmdesigns als Ausgangspunkt für die Operationalisierung und Datenerhebung zu nutzen; die Datensammlung erfolgt dabei im Wesentlichen durch quantitative Erhebungsmethoden, wobei Validität und Reliabilität herausragende Güte-Merkmale des Evaluationsprozesse seien.

Möglichen Konflikten, die aufgrund der unterschiedlichen Bewertung der Evaluationsergebnisse durch die verschiedenen Stakeholdergruppen entstehen, soll der Evaluator mittels 'Konsens-Aufbau', theoretischer Begründung der Programmziele und transparenter Berichterstattung entgegenwirken (vgl. Rossi, Freeman 1982: 311ff; Beywl 1988: 50f). Als problematisch an den Evaluationsmodellen die-

ses Typs wird die unkritische Übernahme der Wertmaßstäbe einflussreicher Stakeholder sowie die Reduktion von Wertfragen auf Input-Output-Relationen angesehen (vgl. Beywl 1988: 51).

Programmziel-/Lernziel-gesteuerte Evaluationsmodelle,
messen die Programmeffekte gegenüber den Programm-/Lernzielen (vgl. Beywl 1988: 45).

Ausgangspunkt dieser Evaluationsmodelle sind die Ziele des Programms bzw. die Lernziele des Curriculums. Das Ergebnis sind dann Diskrepanzaussagen zwischen den Ergebnissen (outcomes) des Programms und den verfolgten Zielen: "Evaluation bestimmt den Wert oder die Angemessenheit eines Objektes auf der Basis der Diskrepanz-Informationen zwischen Standards und Leistung" (Steinmetz 1983: 80).

Die Fragestellungen der Evaluation werden aus den Programmzielsetzungen entnommen und in Hypothesen umgeformt. Das von Carol Weiss entwickelte Modell 'Methods for Assessing Program Effectiveness' nimmt die Effektivitätsmessung mit quantitativen und quasi-experimentellen Methoden vor; nach Ansicht der Entwicklerin ist Evaluation mit empirischer Sozialforschung gleichzusetzen, die Ergebnisse für Entscheidungsprozesse generieren will. Voraussetzung für eine hohe Qualität der Evaluation ist der zweifelsfreie Nachweis des Zusammenhangs von gemessenen Effekten und Programmvariablen, wobei es sinnvoll sei, (wenige) relevante Variablen auszuwählen, statt eine Vielzahl kontrollieren zu wollen. Die hierfür notwendigen Kausalhypothesen sollen aus einem 'theoretischen Begründungszusammenhang' entwickelt werden, wobei in dem Hypothesenzusammenhang die Einflüsse der Kontext- und Störvariablen auf die Effekte bestimmt sein müssen (vgl. Weiss 1972: 73).

Kritisch sind nach Ansicht von Beywl das Vorhandensein eines komplexen Hypothesengerüsts, die kontrollierten Untersuchungspläne sowie standardisierte Untersuchungsinstrumente anzusehen; Voraussetzungen, die nur durch den Ausschluss vieler (situationsspezifischer) Einflussvariablen erreicht werden können (vgl. Beywl 1988: 54ff). Darüber hinaus liege solchen Modellen ein Verständnis von Didaktik zu Grunde, das die Rolle des Lernenden negiert und ihm keinen Einfluss auf den Evaluationsprozess einräume (siehe Heinze u. a. 1981: 13)

Akkreditationsrichtlinien-gesteuerte Evaluationsmodelle,
beurteilen die Evaluationsgegenstände an mehr oder minder expliziten Maßstäben der Anerkennung (vgl. Beywl 1988: 45).

"Akkreditation bezeichnet den Prozess, durch den eine Organisation einer pädagogischen Einrichtung Zulassung gewährt" (Flodan 1983: 261).

Gesteuert wird dieser Zulassungsprozess durch die von der Akkreditationskommission herausgegebenen Richtlinien, die häufig in Zusammenarbeit mit Lehrern und Erziehungswissenschaftlern erstellt werden.

Die Evaluatoren beziehen sich in ihrer Beurteilung einerseits auf Selbsteinschätzungen von Schulen und deren Personal, andererseits auf Untersuchungsprotokolle von Expertenkommissionen (siehe House 1980: 35f). Dabei kommen qualitative, z. T. unsystematische Methoden der Datensammlung zum Einsatz; qualitative Bewertungskriterien können die Erreichung allgemeiner Unterrichtsziele sein (siehe Glass 1972: 181).

Einer der Vorteile dieses Evaluationstypus liege in der schnellen Beurteilbarkeit von pädagogischen Institutionen, was vermutlich auch seine starke Verbreitung begründe (vgl. Beywl 1988: 57f).

Glass kritisiert vor allem die mangelnde Verbreitung objektiver und 'testpsychologischer' Instrumente (vgl. Glass 1972: 181); Stufflebeam/Webster dagegen heben vor allem den Aspekt der 'Korruptierbarkeit' hervor, der sich aus der Möglichkeit der Selbstbeurteilung und den Kommissionsbesuchen ergeben kann (vgl. Stufflebeam/Webster 1981: 76).

Gegnerschafts-gesteuerte Evaluationsmodelle,
sind solche, in denen zwei Parteien – ggf. vertreten durch 'Anwälte' – Nachteile bzw. Vorteile eines Evaluationsgegenstandes gegenüber einer Jury vertreten, die ein Gesamturteil fällen soll (vgl. Beywl 1988: 45).

Ziel dieser wertorientierten Ansätze ist die Evaluation der Wertbezüge der Programmziele und der Bedingungsfaktoren des Programms.

Als methodische Evaluationsvariante hat man das aus der Jurisprudenz bekannte 'Gegnerschafts-Modell' entlehnt, bei dem zwei Evaluationsteams entgegengesetzte Standpunkte (Steuerungskriterium) vertreten, um auf diesem Wege Entscheidungsgrundlagen für eine Jury auszuarbeiten. Diese bewertet dann die Vor- und Nachteile eines Curriculums oder Programms und kommt zu einem abschließenden Urteil (summative Beurteilung; vgl. Beywl 1988: 59f).

Auf Grund festgestellter Defizite (ungleich kompetente Anwalteams u. a.) hat Wolf den ursprünglichen Ansatz, wie er z. B. von Owens vertreten wurde (siehe Owens 1973), umgearbeitet: Neben echten Partizipationschancen aller Beteiligtengruppen hat er vor allem auch die formative Evaluation in den Mittelpunkt gestellt. Ziel ist in seinem Ansatz die Analyse der Wahrnehmungen von Personen, die durch das entstehende oder in der Entwicklung befindliche Programm betroffen sind. Mit diesem Ansatz wird der Anspruch verbunden, dass die Jury als Entscheidungsträger während des Prozesses über die verschiedenen Sichtweisen 'unterrichtet' wird (edukativer Faktor; vgl. Wolf 1983: 192).

Mit dieser Umgestaltung geht nach Ansicht von Beywl auch eine Verlagerung des Steuerungskriteriums einher; waren es ursprünglich die entgegengesetzten Entscheidungsalternativen, sollen nach Wolf die 'Streitpunkte' der relevanten beteiligten Gruppen in den Mittelpunkt des Verfahrens gerückt werden, so dass diese

Verfahren auch den 'Beteiligteninteressen-gesteuerten Evaluationsmodellen' zugeordnet werden könnten (vgl. Beywl 1988: 61).

Stufflebeam/Shrinkfield dagegen sehen in diesen Ansätzen weniger ein Evaluationsdesign als vielmehr eine Methode des "[...] Präsentierens, Interpretierens, Synthetisierens und des Berichtes von Ergebnissen" (Stufflebeam, Shrinkfield 1985: 268).

Kennerschafts-gesteuerte Evaluationsmodelle,

nehmen die auf den professionellen und persönlichen Erfahrungen beruhenden Qualitätsstandards des Evaluators zur Grundlage, die eine Basis für dessen subjektive Urteile über das Programm oder Curriculum bilden (vgl. Beywl 1988: 45)

Nach Eisner sollen Theorien nicht Erklärung und Prognose, sondern Erhellung ('Illumination') bei der Deskription und Bewertung des Evaluationsgegenstandes leisten, so dass die Theorie in Eisners Evaluationsmodell heuristische Funktion inne hat:

"Die Abgabe pädagogischer Werturteile setzt nicht nur voraus, die Feinheiten des Geschehens im Klassenzimmer zu sehen, deren Bedeutung zu interpretieren und deren Funktion bestimmen zu können; hinzukommen muss ein ausreichender Hintergrund an pädagogischer Theorie, Philosophie und Geschichte [...]" (Eisner 1983: 343).

Evaluation weist hiernach zwei Funktionen auf: die 'Kennerschaft' - Verstehen dessen was man erfahren hat und 'Kritik', die als Kunst der Enthüllung bezeichnet wird. Die 'Kritik' wiederum beinhaltet drei Komponenten: Erstens den deskriptiven Aspekt (Beschreibung des Phänomens), zweitens den interpretativen Aspekt (Versuch, den Stellenwert der Handlungsformen für die am pädagogischen Geschehen Beteiligten zu verstehen) und drittens den evaluativen Aspekt (Beurteilung der pädagogischen Wichtigkeit und des Wertes des zu evaluierenden Objekts).

Als methodische Herangehensweise und gleichzeitig als Instrument fungiert der Evaluator, der die Daten sammelt, auswertet, interpretiert und abschließend beurteilt. Gerade hierin manifestiert sich auch die Kritik, denn die 'subjektive' Verarbeitung, d. h. die nicht auf ein Wertemerkmal ausgerichtete Verarbeitung und Deutung von Informationen kann zu einer einseitigen Auslegung und (damit) zu einer parteiischen Stellungnahme führen (vgl. Beywl 1988: 61ff).

Entscheidungs-gesteuerte Evaluationsmodelle,

sind solche, in denen die zur Bewältigung von Bewertungs- und Entscheidungssituationen erforderlichen Informationen zur Verfügung gestellt werden (vgl. Beywl 1988: 45).

Das von Stufflebeam entwickelte CIPP-Evaluationsmodell kann als eines der für diesen Modelltypus charakteristischen Modelle angesehen werden, das den Versuch impliziert, den Fokus nicht mehr auf allgemeine und spezielle Lern- und Pro-

grammziele zu richten, sondern auf solche Entscheidungssituationen, die für die Verbesserung von Curricula richtungsweisend sind:

"Im Allgemeinen bedeutet Evaluation die Gewinnung von Informationen durch formale Mittel wie Kriterien, Messungen und statistische Verfahren mit dem Ziel, eine rationale Grundlage für das Fällern von Urteilen in Entscheidungssituationen zu erhalten" (Stufflebeam 1972: 124).

Die angesprochenen Entscheidungssituationen sind zu Beginn der Evaluation im Hinblick auf Ebene, Schwerpunkt, Funktion, Inhalt, Gegenstand, Zeitpunkt sowie das 'Ausmaß der kritischen Reflektiertheit der Entscheidungen' zu spezifizieren (vgl. Stufflebeam 1972: 131). Der Klassifikation der Entscheidungssituationen sind angemessene Evaluationsarten zuzuordnen, wobei das Ziel, die laufende Verbesserung der pädagogischen Praxis ('improvement-orientated'), unter Nutzung nicht-experimenteller Versuchspläne erreicht werden soll:

- Kontext-Evaluation: Was sind die Ziele des Programms und berücksichtigen diese die Bedürfnisse der Beteiligten?
- Input-Evaluation: Welche Mittel in Bezug auf Pläne, Personal, Budget u. a. sind für das Erreichen der gesetzten Ziele von Bedeutung?
- Prozess-Evaluation: Wie werden die Beteiligten über den Evaluationsprozess informiert? Wie wird die Allokation der Ressourcen vorgenommen?
- Produkt-Evaluation: Welche Belege gibt es für die Ergebnisse des Programms? Soll das Programm aufgegeben, weitergeführt oder verändert werden und ist dazu eine Finanzmittelausweitung oder –verkleinerung notwendig oder möglich?

Darüber hinaus empfiehlt Stufflebeam ein evaluationsbegleitendes, mit Beratungscharakter versehenes, Komitee, das zur Konkretisierung der Fragestellungen und Interpretation der Ergebnisse herangezogen werden soll und sich aus Vertretern der Beteiligtegruppen zusammensetzt; entstehende Probleme sollen durch eine konsensschaffende Methode bereinigt werden (vgl. Stufflebeam, Shrinkfield 1985: 172f).

Mertens fasst die Bemühungen Stufflebeams, die Mängel bestehender Evaluationsansätze zu eliminieren, unter dem Aspekt der Entscheidungsorientierung zusammen, indem sie festhält: "Stufflebeam was instrumental in pushing the definition of evaluation beyond the achievement of objectives to include the idea that it was a process of providing information for decision making" (Mertens 2000: 43).

Aus Teilen des Stufflebeam-CIPP-Modells und durch Kombination mit Elementen der Diskrepanz-Evaluationsmodelle hat Stake sein Modell der responsiven Evaluation entwickelt, das die Idee des Vergleichs von beobachtbaren mit erwarteten

Werten (siehe Fragestellungen zu Evaluationsbereichen im CIPP-Modell) aufgreift.²⁸

Als kritische Aspekte können im Hinblick auf das CIPP-Modell die Unveränderbarkeit der allgemeinen und spezifischen Ziele auf Ebene der Vorplanungsphase, das Unvermögen der quantitativen Methoden neue Ziele zu bestimmen und auf situationsbedingte Besonderheiten zu reagieren sowie einseitige Festlegung bei der Zielformulierung durch Projektträger oder Geldgeber, während die Bedürfnisse der Programmteilnehmer nur als Informationen in die Kontextevaluation einfließen, bezeichnet werden (vgl. Beywl 1988: 67f). Cronbach versucht diesen Kritikpunkten durch Zuweisung einer edukativen Rolle an den Evaluator entgegenzuwirken, dessen Erfolg zumindest teilweise daran gemessen wird, wie gut er die Kommunikation zwischen den Programmbeteiligten herzustellen und Lernprozesse zu initiieren vermag (vgl. Cronbach 1973).

Die genannten Kritikpunkte am CIPP-Modell sowie Erkenntnisse von Cronbach hat Stufflebeam in Zusammenarbeit mit Shrinkfield dann auch zur Weiterentwicklung des entscheidungsgesteuerten Evaluationsmodells genutzt:

Konsumentenbedürfnis-gesteuerte Evaluationsmodelle,
beurteilen vielfältige Merkmale des Evaluationsgegenstandes an vom Evaluator eingeschätzten Bedürfnissen der Konsumenten und Klienten (vgl. Beywl 1988: 45).

Beywl fasst unter diesem Modelltyp verschiedene Ansätze von Scriven zusammen, in deren Mittelpunkt die 'Bedürfnisse' der Beteiligten stehen und als sog. 'Steuerungskriterien' fungieren. Gegenstand der Evaluation ist daher auch nicht allein das Erreichen von Programm- oder Lernzielen zu prüfen, sondern die Ziele selbst (vgl. Scriven 1972: 72), so wie es im 'goal-free-Modell' zum Ausdruck kommt, werden Evaluationsgegenstände (Scriven 1973).

Scriven unterscheidet zwischen Zielen des Programms und solchen, die unabhängig von Programm identifiziert werden können. 'Needs' sind nach Scriven alles "[...] was essentiell für eine befriedigende Lebensweise ist, alles, ohne das diese Lebensweise oder das Leistungsniveau unter ein befriedigendes Level fallen würde" (Scriven, zit. n. Stufflebeam, Shrinkfield 1985: 318).

Demnach untersucht der 'zielfreie' Evaluator nicht nur gestellte Hypothesen oder Programm-/ Lernziele, sondern sucht nach allen Ergebnissen (outcomes) des Programms. Nach Ansicht von Scriven beschreiben gerade die nicht intendierten Effekte den Wert des Programms (vgl. Scriven 1973: 321).

Die Kritik an diesem Ansatz ergibt sich vor allem aus dem nicht vorhandenen expliziten Ablaufschema; Scriven hat lediglich eine 'Key evaluation checklist' zusammengestellt, die in 15 Punkten alle für ein 'gutes' Evaluationsdesign notwendigen Aspekte umfasst. Durch diese bewusst herbeigeführte Offenheit besteht die

28) Mertens sieht in der Fortentwicklung der Evaluationsmodelle generell den Versuch, Mängel in bestehenden Evaluationsdesigns zu vermeiden (vgl. Mertens 2000: 43).

Gefahr der einseitigen Auslegung und damit der generellen Verzerrung, des Missbrauchs des wertbezogenen Ansatzes (vgl. Beywl 1988: 75).

Beteiligteninteressen-gesteuerte Evaluationsmodelle, stellen die Anliegen und Konfliktthemen der Programmbeteiligten in den Mittelpunkt, diese geben dem Evaluator Handlungsimpulse für die Problemdefinition, die Datensammlung, die Interpretation und die Beurteilung (vgl. Beywl 1988: 45).

Unter diesem Typus lassen sich eine große Anzahl von Modellen subsumieren, die unter verschiedenen Begriffen in der Literatur aufzufinden sind (Guba – naturalistische Evaluation, Stake – Fallstudienansatz u. a.).

Übereinstimmung erreichen die fallstudienorientierten Ansätze im Ansatzpunkt der Evaluationstätigkeit: sie alle betonen die 'Selbstverantwortung' der lokalen Beteiligengruppen, deren Anliegen in den Mittelpunkt der Betrachtung gerückt werden.

Der Evaluator greift die geäußerten Interessen auf und verfolgt diese im Rahmen des Evaluationsprozesses (katalytische Funktion; siehe auch House 1983). Die Grundlage für die Interpretation der Ergebnisse sind dann auch die unterschiedlichen Wert-Perspektiven der Beteiligengruppen.

"Induktivität, Untersuchung in naturalistischer Situation und ganzheitliche Sichtweise, kombiniert mit gezielter Beteiligung der Betroffenen und Akteure zum Zwecke einer Demokratisierung der Evaluation unterscheiden diesen Modelltyp [...] (Beywl 1988: 77)" nach Ansicht von Beywl von anderen Evaluationstypen.

Problematisch bei Modellen dieses Typs ist die Frage der Gewichtung der zu verfolgenden Interessen, so dass auch hier die Gefahr der Verzerrung auf Grund von Machtpolaritäten besteht.

Ungeachtet der spezifischen Ausrichtungen der verschiedenen Evaluationsansätze, ihrer immanenten Vor- und Nachteile sowie ihrer Einordnung in die Evaluationstypologie(n) ist der Literatur eine die ganze methodologische Diskussion umfassende Kritik zu entnehmen:

Nach Lakatos ist der rationale Vergleich alternativer Theorien der zentrale Gedanke der Methodologie (vgl. Lakatos 1974: 114). In Bezug auf die Evaluationsmethodologie hat jedoch gerade 'Theorielosigkeit' Bickman 1987, Chen 1990, Chen/Rossi 1980, Palumbo/Gregware 1992 u. a. veranlasst, programmtheoretische Ansätze zu entwickeln, um dem Prozess der Weiterentwicklung der 'atheoretischen' Evaluation entgegenzuwirken. Der in der fachliterarischen Diskussion häufig vertretene Ansatz von Chen wird nachfolgend erläutert, um beispielhaft die Implikationen einer theoriegeleiteten Evaluation darzustellen. Stufflebeams Einschätzung programmtheoretischer Evaluationsansätze schließt diesen Diskussionsschwerpunkt in der Evaluationsforschung ab und leitet zur zweiten Zwischenbilanz unter dem Fokus des Standes der Literatur in den Themenschwerpunkten 'Evaluationsdefinition, Methodologiediskussion und programmtheoretische Evaluation' über.

4.3 Theoriegeleitete Evaluation

4.3.1 Aufbau und Dimensionen des programmtheoretischen Ansatzes von Chen

Die Präferenz von 'input/output'-bezogenen Evaluationsmodellen sowie die 'Antipathie' gegenüber der Theorieorientierung hat Wortmann dazu geführt (Programm-) Evaluation als 'multi-disziplinäre' und weitestgehend atheoretische Aktivität zu beschreiben (vgl. Wortmann 1983: 224).

Unter Missachtung von 'Theorie als Referenzrahmen', der das Verstehen von Forschungsbefunden ermöglicht und Richtlinien für die Analyse eines 'Phänomens' liefert (vgl. Chen 1990: 17; siehe auch Rudner 1966), sind lediglich 'black-box'-Evaluationen (siehe hierzu u. a. Lipsey 1987) unter Vernachlässigung der kausalen Mechanismen des Programmes (des Modellversuchs) durchführbar, so dass Mängel nicht identifiziert und Beziehungen zwischen geplanten und tatsächlichen Interventionen sowie beabsichtigten und nicht beabsichtigten Ergebnissen nicht hergestellt werden können (vgl. Chen 1990: 18).

Die einseitige Ausrichtung auf methodische Aspekte in der Evaluationsforschung und die sich regelmäßig wiederholende Betonung von Stärken und Schwächen einzelner Evaluationsansätze bzw. Typen sowie z. B. die Diskussion quantitative vs. qualitative 'Forschungsmethoden' führt jedoch zu einer Fokusverengung. Das Ergebnis ist die Generierung einer kontinuierlichen Debatte und die Polarisierung innerhalb der Forschungsgemeinschaft. Von Bedeutung jedoch ist die Entwicklung theoretischer Ansätze, die die systematische Integration kausaler Faktoren in Evaluationsprozesse vorantreiben.

In Bezug auf die Identifikation der relevanten Aspekte eines zu evaluierenden Programms fordert u. a. Trochim die kausalen Mechanismen im Rahmen einer 'Programmtheorie' zu prüfen, um einer Verfestigung der theoriefremden und kontextlosen Ansätze entgegenzuwirken (vgl. Trochim 1986: 3). Chen formuliert hierzu:

"Program theory can provide guidelines for identifying which issues are most important in an evaluation, determining what method or methods are most relevant to address these issues, and suggesting how to apply the best method or methods or dealing with these issues" (Chen 1990: 28).

Die intensiven Debatten zwischen den verschiedenen Vertretern der Evaluationstypen hätten gezeigt, so Chen, dass es keine Knappheit an Methoden in der Evaluation gäbe. Vielmehr zeige die Unzufriedenheit mit den Entwicklungen in der Evaluationsforschung, dass die Notwendigkeit für eine theoretische Perspektive bestehe, die umfassend genug ist, nicht nur Aspekte der Programmimplementation, der kausalen Prozesse, des Programmdesigns und der Programmresultate zu erfassen, sondern auch ausgereift ist, Lenkung anzubieten im Umgang mit multiplen oder sogar konfliktären Meinungen und Erkenntnissen (vgl. Chen 1990: 34).

Chen geht in seinem programmtheoretischen Ansatz (siehe Abb. 10) von den Prämissen aus, dass ein Programm immer eine Form von Theorie beinhalte, auch dann, wenn diese Theorie nicht expliziert wird und auf keine systematische Grund-

lage zurückzuführen ist. Darüber hinaus sei eine theorieorientierte Perspektive in Evaluationsprozesse zu integrieren und die Art der Theorie zu prüfen.

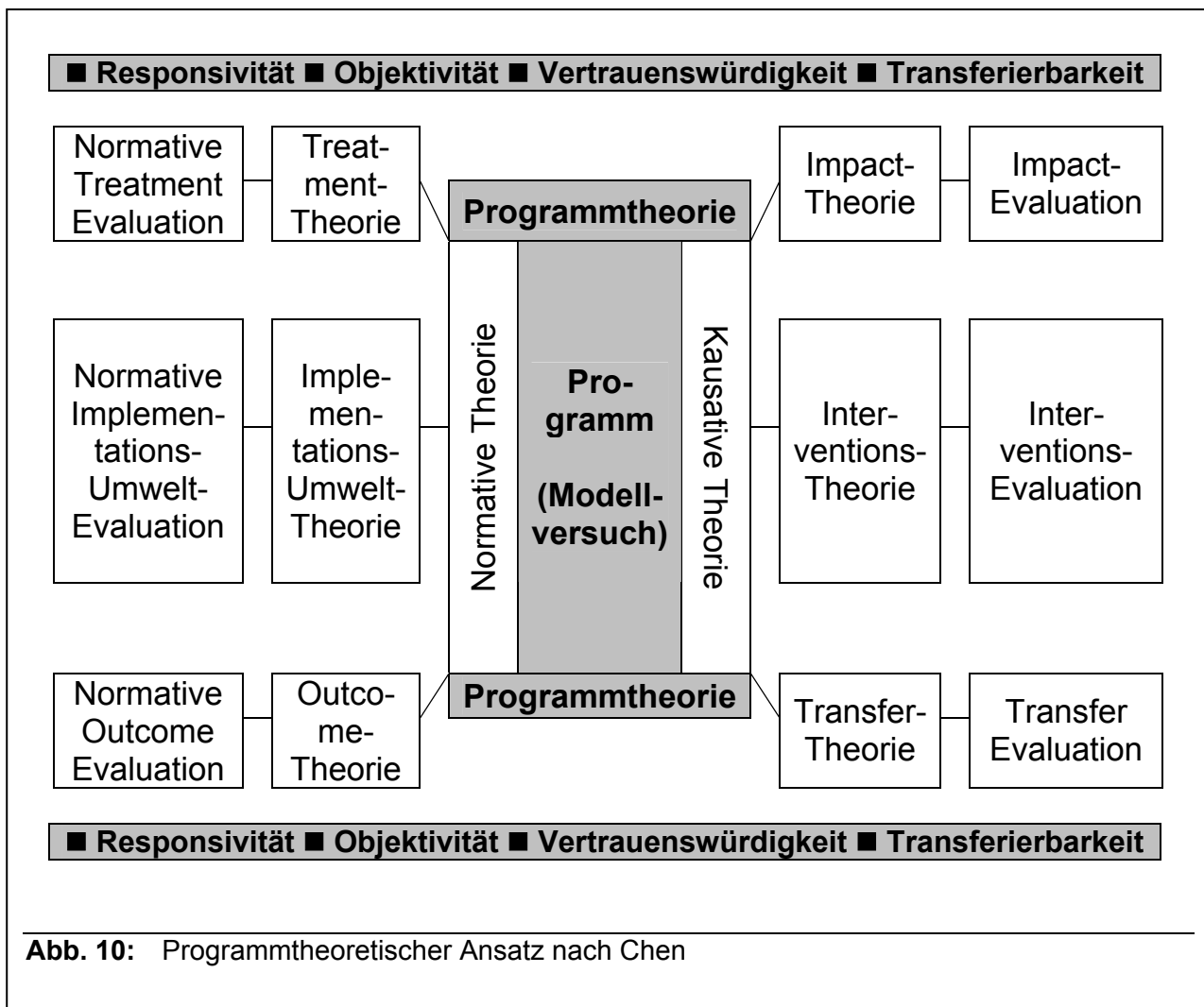


Abb. 10: Programmtheoretischer Ansatz nach Chen

Im Gegensatz zu den Sozialwissenschaften, in denen im Allgemeinen 'Theorie' ein deskriptiver Charakter zugeschrieben werde, um ein Phänomen zu beschreiben und zu erklären, ohne jedoch handlungsanleitend zu wirken, definiert Chen 'Theorie' als ein Satz von einander bedingender Annahmen, Prinzipien und/oder Vorschlägen, um soziale Handlungen zu erklären oder anzuleiten (vgl. Chen 1990: 39f).

Während Bickman, Lipsey und Wholey auch für die Programmtheorie die Bedeutung deskriptiver Funktion hervorheben (vgl. Bickman 1987: 5; Lipsey 1987: 7; Wholey 1987: 78) betont Chen, dass Programmtheorie bei der Auswahl der Kriterien, um die Programmausprägungen zu beurteilen ebenso Werturteile hierzu einschließe und außerdem Aussagen darüber tätige, wie ein Programmdesign zu gestalten und umzusetzen ist.

Hiernach beinhalte die Programmtheorie sowohl deskriptive als präskriptive Elemente und sei durch folgende (wertorientierte) Aspekte charakterisiert:

- Handlungsanleitung (action orientation): Programmtheorie beinhaltet Strategien zur Erreichung gesetzter Ziele oder dem Beheben sozialer Probleme. Sie impliziert die Notwendigkeit der Auseinandersetzung/der Handlung, um eine Situation zu verbessern.
- Programmkonzeption und Implementationsstrategie (Conceptualization of Treatment Format and Implementation Strategy): Aus der Handlungsanleitung resultiert die Notwendigkeit von Angaben über Aufbau und Umsetzung des Programmdesigns sowie das Profil derjenigen Personen, die an der Programmdurchführung beteiligt sind. Gegenstand der Programmtheorie sind weiterhin die Vorgaben der Auftraggeber, die in das Programm aufzunehmen sind. Außerdem sind die dem Programmdesign zugrunde liegenden Prämissen offen zu legen, um diese untersuchen und so ein Feedback für spätere Entscheidungszwecke geben zu können.
- Ergebnis-, Messkriterien (Range of Options in Choosing Outcome Criteria): Die Interessen und Werte eines Programms hängen von der 'Systemposition' der Stakeholder ab, so dass diese in Art und Ausprägung sehr unterschiedlich sein können. Generell kann zwischen 'Erhaltungs- und Entwicklungszielen' unterschieden werden kann. Aufgrund von Ressourcenbeschränkungen ist auf jeden Fall eine Selektion der Ergebniskriterien vorzunehmen, was den präskriptiven Aspekt der Programmtheorie noch einmal betont.

Insgesamt gelte, dass in der Kombination der drei wertorientierten Merkmale und unter Berücksichtigung der kausalen Prozesse die duale Ausprägung der Programmtheorie (deskriptiver und präskriptiver Charakter) hervortrete (vgl. Chen 1990: 40ff). Diese Dualität von (Programm-) Theorie sei es, die einerseits wissenschaftliche Leistungen hervorbringe und das Potenzial beinhalte, die Gesellschaft zu verändern (vgl. Dewey 1929 u. 1933).

Für das von Chen entwickelte Modell der Programmtheorie gilt auf Grundlage der dargestellten drei Merkmale folgende Definition:

Program theory is "[...] a specification of what must be done to achieve the desired goals, what other important impacts may also be anticipated, and how these goals and impacts would be generated" (Chen 1990: 43).

Diese Definition betont zwei Funktionen einer Programmtheorie. Zum einen werden Aussagen gefordert, wie die Struktur des Programms gestaltet sein soll, also Aspekte wie Programmdesign, Auswahl der Programmziele sowie Durchführung von Implementationsprozessen, inklusive Offenlegung der Grundannahmen und –werte des Programms (präskriptiver Theorieteil); der zweite Teil dagegen zielt auf die Darstellung der kausalen Beziehungen zwischen Programmdesign, Implementationsprozessen und Ergebnissen ab (deskriptiver Theorieteil).

Der Teil der Programmtheorie, der sich auf die präskriptive Theorie bezieht, wird nachfolgend in Bezug auf Chen als 'normative' Theorie bezeichnet; die deskriptive

Theorie beschreibt Chen als 'kausative Theorie'. Beide Theorien seien Subtheorien der Programmtheorie.

Während die normative Theorie Anleitung bietet im Hinblick auf die Auswahl der Ziele und zu erreichenden Ergebnisse sowie Informationen über den Aufbau und die Umsetzung des Programmdesigns, spezifiziert die kausative Theorie die Arbeitsweise des Programms, indem sie die Bedingungen identifiziert, unter denen bestimmte Prozesse entstehen und welches die wahrscheinlichen Konsequenzen hieraus sein werden (vgl. Chen 1990: 43ff).

Die beiden Subtheorien werden weiterhin in sechs Theorie-Dimensionen unterteilt, denen je ein Evaluationstyp aus der von Chen aufgezeigten 'Typologie theoriegesteuerter Evaluationen' gegenübergestellt wird²⁹ (siehe Abb. 10):

Die Treatment-Theorie (n. T.)³⁰ wird als Basiselement innerhalb der Programmtheorie angesehen, die die intendierten Änderungen innerhalb des untersuchten Programms, Curriculums u. ä. ermöglicht. Sie beschäftigt sich mit der Konzept-/Designentwicklung der Programm- oder Curriculumstruktur. Wichtig sei die Generierung einer Struktur, die ex-ante eine systematische Evaluation der Prozesse und Ergebnisse berücksichtigt. Darüber hinaus müssten Strategien entwickelt werden, die Abweichungen von der geplanten Umsetzung aufzuzeigen helfen (vgl. Chen 1990: 46).

Dieser Theoriedimension weist Chen den Typus der 'normativen Treatmentevaluation' zu, welche darauf abzielt, die normative Struktur des Treatments zu identifizieren, die realisierte Umsetzung im Programm zu untersuchen und die Kongruenz zwischen geplanter und durchgeführter Umsetzung zu beurteilen (vgl. Chen 1990: 52f).

Gegenstand der Implementations-Umwelt-Theorie (n. T.) ist die Beschäftigung mit der Beschaffenheit der Implementationsumgebung (-umwelt). Informationen hierüber können dafür genutzt werden, die Umsetzungsprozesse des Programms zu verbessern und die Interpretation der Evaluationsergebnisse zu erleichtern. Hierdurch werden Aussagen möglich, die sich auf die wünschenswerten Profile der räumlichen, personalen und curricularen Gegebenheiten beziehen (vgl. Chen 1990: 46f).

Der Typus der 'normativen Implementations-Umwelt-Evaluation' untersucht und beurteilt die Kongruenz zwischen der in der Theorie-Dimension beschriebenen Implementationsumgebung und der tatsächlichen Implementationsumgebung; hierzu sollen empirische Daten gesammelt werden (vgl. Chen 1990: 53f).

Die Outcome-Theorie (n. T) setzt sich mit Aspekten des Umgangs mit beabsichtigten und unbeabsichtigten Ergebnissen auseinander. Die Notwendigkeit ergibt

29) Auf die Möglichkeit der Kombination verschiedener Evaluationstypen wird nicht eingegangen, da diese zu vielfältig sind.

30) Für die Darstellung der Theoriedimensionen werden zwecks Einordnung in die beiden Subtheorien die Abkürzungen n. T. (normative Theorie) und k. T. (kausative Theorie) verwendet.

sich aus dem Zweck eines Programms oder Curriculums, vorher festgelegte Ziele oder beabsichtigte Ergebnisse erreichen zu wollen, die als Kriterien verwendet werden, die Wirksamkeit eines Programms/Curriculums zu beurteilen (vgl. Chen 1990: 47).

Die 'normativen Outcomeevaluation' setzt einen Satz von Programmzielen voraus, um den Prozess der Programmplanung zu erleichtern. Gerade bei einer Vielzahl unterschiedlicher Stakeholder sei dieser Evaluationstypus sinnvoll, um zu klären, welche Ziele verfolgt werden sollen bzw. um divergierende Zielsetzungen einander annähern zu lassen (vgl. Chen 1990: 54).

Der Impact-Theorie (k. T.) obliegt die Auseinandersetzung mit der Einwirkung des Treatments (Programm, Curriculum) auf das Ergebnis (Outcome). Ziel ist es, begründete Schlussfolgerungen in Bezug auf den Einfluss der Einwirkungen (z. B. Curriculum) auf die Ergebnisse zu ziehen, da die Stakeholder Informationen hinsichtlich der Effektivität des Programms fordern (vgl. Chen 1990: 47f).

Die 'Impact-Evaluation' unterscheidet sich von der traditionellen summativen Evaluation in zwei Punkten: Erstens hebt die Impact-Evaluation das Generieren einer breiten Beweisgrundlage für das Beurteilen von Programmwirkungen hervor. Um diese zu erhalten, sind Evaluationen erforderlich, die 'aktiv und kreativ' Strategien und Pläne nutzen, vor allem solche, die theoriegeleitet sind, um Belege für Schlussfolgerungen in Bezug auf kausale Zusammenhänge zu liefern. Zweitens betont dieser Evaluationstypus die Notwendigkeit der Sensibilität gegenüber beabsichtigten und unbeabsichtigten Ergebnissen (Outcome; vgl. Chen 1990: 54).

Untersuchungsaspekte der Interventions-Theorie (k. T.) sind die kausalen Prozesse zwischen dem implementierten Treatment und den Ergebnissen (Outcome). Eine Untersuchung der intervenierenden Variablen erzeugt Informationen über die Funktionsweise des Programms und hilft die Stärken und Schwächen zu identifizieren, so dass Verbesserungen möglich werden (vgl. Chen 1990: 48).

Der Blickwinkel der Evaluation wird im Typus der 'Interventions-Evaluation' (k. T.) durch die intervenierenden Prozesse/Variablen erweitert. Der kausale Prozess, der durch diesen Evaluationstyp identifiziert wird, liefert nicht nur Informationen, die wichtig sind, um das Potenzial eines Programms besser zu verstehen, sondern liefert auch eine Diagnose der potenziellen Probleme, die genutzt werden kann für zukünftige Programmverbesserungen (vgl. Chen 1990: 54f).

Die Transfertheorie (k. T.) behandelt Fragen der Übertragung von Evaluationsergebnissen aus dem 'Forschungssystem' in ein zukünftiges Zielsystem. Aufgrund der nicht zwingend vorhandenen Homogenität der beiden Systeme sind weitere Bemühungen notwendig, die Transferfähigkeit zu verbessern (vgl. Chen 1990: 48f). Diese Art der Evaluation (Transferevaluation) ist durch eine prospektive Ausrichtung gekennzeichnet, mit dem Ziel, die Transferfähigkeit des Evaluationsgegenstandes zu optimieren. Chen und Rossi betonen, dass bei der Gestaltung der Trans-

der Evaluation Strukturen und Arbeitsweisen der zukünftigen Organisationen, d. h. der Implementationsumgebung, berücksichtigt werden müssen (vgl. Chen 1990: 55).

Die Dimensionen der normativen und kausativen Subtheorien seien nicht getrennt voneinander zu betrachten. Vielmehr würde die Evaluation normativer Theoriedimensionen Informationen hinsichtlich der kontextuellen und einflussnehmenden Faktoren erzeugen, die erforderlich sind, um die empirischen Erkenntnisse aus den kausativen Dimensionen wie Wirkung, intervenierende Variable und Transfer, zu erklären und zu interpretieren. Hierdurch würden Erkenntnisse über die 'dazwischen liegenden Prozesse' ermöglicht, die darauf schließen lassen, wie diese ablaufen und in welchem Maße sie ähnliche oder gleiche Ergebnisse in unterschiedlichen Programmen (Modellversuchen) oder Kontexten produzieren können.

Durch die Kombination aller sechs Theoriedimensionen werde eine übergeordnete 'Theorie' des Programms erzeugt und es sei wünschenswert, dass alle sechs Theorien evaluiert würden, was jedoch aus Gründen knapper Ressourcen oder besonderer Interessen der Stakeholder nicht immer möglich sei (vgl. Chen 1990: 51f).

4.3.2 Werteorientierung im programmtheoretischen Ansatz von Chen

Der Entwurf einer theoriegesteuerten Evaluation erhält im Wesentlichen sein Profil durch die 'Werte' ("Values are defined [...] as ideas or meanings of what ought to be"; Chen 1990: 57), die mit ihr verbunden werden. Durch sie wird definiert, welche Ziele oder Ergebnisse evaluiert werden, welche kausalen Prozesse untersucht und wie die betroffenen Variablen konzeptualisiert und gemessen werden sollen.

Ein in der Literatur ausführlich diskutiertes Problem ist die Auswahl der 'gültigen' Werte (Gütemerkmale; siehe hierzu u. a. Lamnek 1995: 3ff); die Vertreter des quantitativen Forschungsparadigma heben vor allem die Bedeutung interner und externer Validität für die Evaluation hervor, wobei auch in diesem Zusammenhang kein Konsens hinsichtlich des Stellenwertes der beiden 'Gütekriterien' besteht. Während z. B. Campbell, Stanley und Koch (1963, 1979) der internen Validität Priorität gegenüber der externen Validität einräumen, da das Erlangen der internen Validität die Voraussetzung für die externe Validität sei, heben Kritiker wie Cronbach (1980, 1982) die umgekehrte Prioritätenreihenfolge mit der Begründung hervor, dass eine zweckerfüllende Evaluation die Aufmerksamkeit der Stakeholder auf entscheidungsrelevante Informationen lenke und dass diese den Evaluator in die Lage versetzten, die Entwicklung relevanter Projekt-Dimensionen zu extrapolieren.

Lincoln und Guba (siehe Guba 1989) haben eine Klassifizierung von Werten entwickelt, die im Wesentlichen auf dem qualitativen Forschungsparadigma basieren und Gütemerkmale wie Echtheit der Werte, Anwendbarkeit, Konsistenz und Neutralität betonen.

Chen stellt 'die Werte' der quantitativen und qualitativen Forschungsparadigmata einander gegenüber und konstatiert weitreichende Übereinstimmung:

Qualitatives Forschungsparadigma	< ≈ >	Quantitatives Forschungsparadigma
▪ Echtheit der Werte	< ≈ >	▪ interne Validität
▪ Anwendbarkeit	< ≈ >	▪ externe Validität
▪ Konsistenz	< ≈ >	▪ Reliabilität
▪ Neutralität	< ≈ >	▪ Objektivität

Abb. 11: Gegenüberstellung der 'Werte' des quantitativen und des qualitativen Forschungsparadigmata

Differenzen zwischen diesen Sätzen von Güte Merkmalen seien vor allem in Bezug auf deren Erreichung zu sehen, weniger in deren inhaltlicher Bedeutung/Ausrichtung. Daher müsse der Fokus der Diskussion hauptsächlich auf die Bedeutung der Werte selbst gelegt werden und weniger auf die Frage, mit welchen Methoden diese Werte implementiert werden können. Geschehe dieses, dann seien keine zwei Bezeichnungen für den gleichen Wert, einerseits aus quantitativer und andererseits aus qualitativer Sicht, erforderlich.

Für eine weiterführende generelle Diskussion hinsichtlich der Beziehung zwischen Werten und anderen Aspekten der Evaluation sei vor allem ein übersichtliches Schema von Evaluationswerten (Gütemerkmalen) erforderlich. Durch das Zusammenfassen zweier oder mehrerer Werte gleicher Zielrichtung ergibt sich für Chen folgendes Set von Werten/Gütemerkmalen (vgl. Chen 1990: 58ff):

- | | |
|------------------|------------------------|
| ▪ Responsiveness | ▪ Vertrauenswürdigkeit |
| ▪ Objektivität | ▪ Transferierbarkeit |

▪ Responsiveness

Abweichend vom traditionellen Begriff der Evaluation, bei dem nur die Interessen einer Stakeholdergruppe berücksichtigt werden (Entscheidungsträger), ist eine Evaluation, die das Güte Merkmal der 'Responsiveness' anerkennt vor allem gekennzeichnet durch die Pluralität der Interessenorientierungen, um so die Nützlichkeit der Evaluation zu steigern (siehe Cronbach 1980). Aspekte, die 'Responsiveness' darüber hinaus als Güte Merkmal charakterisieren sind Rechtzeitigkeit der Informationsbereitstellung, um deren Nutzung zu gewährleisten, Relevanz für Entscheidungsfindungsprozesse und Weite, damit die ganze Bandbreite der beabsichtigten und nicht beabsichtigten Ergebnisse aufgenommen werden können.

▪ Objektivität

Scriven stellt die Begriffe 'Subjektivität' und Objektivität zwecks Abgrenzung gegenüber: "'Subjective' means unreliable, biased or probably biased, a matter of opinion, and 'objective' means reliable, factual, confirmable or confirmed, and so forth" (Scriven 1971: 95f). Darüber hinaus werde 'Objektivität' eher durch die intersubjektive Vereinbarungen über Werte erreicht, als dadurch, dass alle Stakeholder ihre Werte unabhängig voneinander verdeutlichen. Auf Grundlage der Ausführungen Scrivens verbindet Chen 'Objektivität' mit der Erwartung, dass Auswertungsergebnisse nicht durch persönliche Präferenzen der Evaluatoren beeinflusst werden,

so dass Resultate, die durch bestimmte Forschungsmethoden erzielt werden, durch andere Evaluatoren mit denselben Methoden wiederholbar sind.

▪ **Vertrauenswürdigkeit**

Dieses Gütemerkmal definiert Chen als die Versicherung, dass eine Evaluation überzeugende Belege liefern wird, denen die Stakeholder vertrauen können. Eine Verbindung sei zur 'internen Validität' im quantitativen Paradigma zu sehen, da auch hier das Hauptbestreben in der Eliminierung verfälschender Faktoren liege. Auch Guba und Lincoln, als Vertreter des qualitativen Paradigmas, seien zu der Überzeugung gekommen, dass der von ihnen beschriebene Gütestandard 'wahrer Wert' inhaltlich mit dem der 'Vertrauenswürdigkeit' übereinstimme (siehe Guba, Lincoln 1989).

▪ **Transferierbarkeit**

Der Transferaspekt bezieht sich auf den Umfang in dem die Evaluationsergebnisse auf zukünftige Probleme angewandt werden können. Er impliziert somit, dass Evaluationsergebnisse, die in vorangegangenen Programmaktivitäten gesammelt worden sind, von Stakeholdern zur Entscheidungsfindung in der Zukunft genutzt werden. Wenn Evaluatoren jedoch keine Informationen über die Hintergründe eines Programms liefern, so dass die Entscheidungsträger die Evaluationsergebnisse nicht auf zukünftige Bereiche extrapolieren können, dann habe eine Evaluation nur einen geringen 'Wert' (vgl. Chen 1990: 61ff).

Je nach Präferenz im Hinblick auf die unterschiedlichen Gütemerkmale/Werte ist es in der Literatur zur Entwicklung verschiedener Programmtheorie-Ansätze gekommen. Chen greift in seinen Ausführungen den 'Stakeholder-Ansatz' und den 'sozialwissenschaftlichen Ansatz' auf, um aus der Diskussion der Stärken und Schwächen einen integrativen Programmtheorieansatz aufzuzeigen, der den Versuch unternimmt, Aspekte beider 'Zielrichtungen' aufzunehmen, um so die von ihm hervorgehobenen Gütemerkmale (s. o.) berücksichtigen zu können. Dabei kommt der 'Responsivität' eine integrative Funktion zu:

"Any integrative approach that incorporates the value of responsiveness needs dual theorizing processes to integrate the social science and stakeholder approaches. On one hand, the evaluator should use the stakeholder approach to clarify and refine the key stakeholders' theory through interviews and documentation review [...]. On the other hand, based upon the social science approach, the evaluators should apply their knowledge, expertise, and understanding of the program to construct an alternative theory" (Chen 1990: 71).

Während der Stakeholder-Ansatz vor allem auf die Beurteilung der 'Machbarkeit' eines Programms abstellt, um die Programmleistung zu verbessern und dabei die Interessen der Stakeholder mit Schlüsselfunktionen (Responsivität), wie Finanzierung, Management u. a. reflektiert, betont die sozialwissenschaftliche Variante hauptsächlich die 'Objektivität' in der Konstruktion der Programmtheorie und verpflichtet daher zur Berücksichtigung existierender sozialwissenschaftlicher Theo-

rie(n) ebenso wie der Aufnahme von Informationen über den tatsächlichen Verlauf von Programmen. Hierdurch werde ein Einblick in Programmstrukturen und -prozesse sowie Konsequenzen hieraus ermöglicht (vgl. Chen 1990: 65ff).

Für den programmtheoretischen Ansatz von Chen bildet zunächst der sozialwissenschaftliche Ansatz die Basis. Dieser wird erweitert durch die Gütemerkmale 'Vertrauenswürdigkeit' und 'Transfer'. Die 'Vertrauenswürdigkeit' einer Evaluation könne durch die Einbindung/die Integration in eine Programmtheorie gesteigert werden, da diese theoretische Muster vorgebe, auf die hinsichtlich der Einschätzung des Programms Bezug genommen werden kann. Darüber hinaus müsse eine 'transferorientierte Programmtheorie' die Implementationsumwelt extrapolieren und diese bei der Programmkonzeption berücksichtigen.

Für die Entwicklung einer Programmtheorie schlägt Chen (auch in Zusammenarbeit mit Rossi) folgende Richtlinien vor:

1. Spezifikation der Bedingungs- und Interaktionsvariablen sowie der intervenierenden Variable, die die kausale Beziehung zwischen Treatmentvariablen und Outcomevariablen sowohl im 'Forschungssystem' (Explikationsmodellversuch) als auch im 'Transfersystem' (Planungsmodellversuch) beschreiben.

Rossi und Freeman stellen in diesem Zusammenhang auf die Entwicklung eines Interventions- und Wirkungsmodells ab, das versucht, Vorstellungen hinsichtlich der Regulierung, Modifikation und Kontrolle sozialen Verhaltens oder sozialer Bedingungen in Hypothesen zu übersetzen, auf die als Basis für zukünftige Maßnahmen zurückgegriffen werden kann. Während das Interventionsmodell die Funktionsweise eines Programms zu erklären versucht, enthält das Wirkungsmodell Aussagen über die potenziellen Beziehungen zwischen dem Programm und seinen Zielen sowie eine Strategie hinsichtlich der Annäherung (Erreichung) von Planungszielen und aktuellen Verhaltensweisen oder Programmbedingungen.

Die Kausalhypothese stellt den Kern des Wirkungsmodells dar und ist eine (hypothetische) Aussage über die Wirkung einer oder mehrerer Variablen auf Verhaltensweisen oder Bedingungen, die durch das Programm geändert werden sollen. Interventionshypothese: "Dieser Hypothesentyp bezieht sich auf die Wirkung des Programms bzw. der Intervention auf die in der Kausalhypothese genannten Ursachen" (Rossi, Freeman 1988: 26). Kann eine Veränderung dieser Ursachen vorgenommen werden, dann wird vermutet, dass dies Auswirkungen auf die 'abhängige Variable', d. h. auf die zu beeinflussenden Faktoren, hat. Die Aktionshypothese dient der Begründung, warum die Veränderung der 'unabhängigen Variablen' zu einer Modifikation der zu verändernden Zustände oder Verhaltensweisen führt (vgl. Rossi, Freeman 1988: 25f).

2. Entwicklung eines Verständnisses der Funktionsweise des Programms unter Einbeziehung der betroffenen Stakeholdergruppen und den Stellen, denen die Verantwortung für die Durchführung des Programms übertragen wurde, sowie Berücksichtigung von alternativen Programmentwicklungen und -interaktionen.

3. Generierung einer 'Musterstrategie' und eines 'administrativen Apparates der Evaluation', um so, möglichst präzise, das Profil der zukünftigen Programmbedingungen (siehe Punkt 2) wiederzuspiegeln (vgl. Chen 1990: 69f).

Hinsichtlich der einzusetzenden Evaluationsmethoden stellt Chen auf eine Methodenvielfalt ab (vgl. Chen 1990: 143ff): Während der Stakeholder-Ansatz tendenziell qualitative Evaluationsmethoden zum Einsatz kommen lässt, bezieht der sozialwissenschaftliche Ansatz vor allem auch quantitative Erhebungsmethoden ein. Durch die Integration beider Ansätze kommt es zu einer Methodenpluralität in der Evaluation der Programmtheorie. Rossi und Freeman betonen im Zusammenhang mit der Diskussion um die Stärken und Schwächen verschiedener Evaluationsdesigns, dass es nicht die 'perfekte' Evaluation gäbe, sondern lediglich ein 'optimales Design', mit dem die Programmziele in 'genügendem' Umfang ermittelt werden können (vgl. Rossi, Freeman 1988: 94).

Im Folgenden Abschnitt soll der Frage nachgegangen werden, welche Stärken und Schwächen die Gruppe/der Typus der programmtheorieorientierten Evaluationsansätze aufweisen. Hierzu dient eine Untersuchung von Stufflebeam - als Ergebnis werden Aussagen erwartet, inwieweit die Evaluationsforschung Ansätze diskutiert/generiert hat, die eine theoriegeleitete Evaluation ermöglichen und ob diese 'hinreichende' Informationen erzeugen und somit den Anforderungen von Lakatos/Wortmann genügen

4.4 Klassifikation und Analyse von theoriebasierenden Programmevaluationsansätzen nach Stufflebeam

4.4.1 Darstellung des Evaluationsvorhabens

Stufflebeam untersucht 22 Programmevaluationsansätze, die seit etwa 1960 (bis 1999) in der Literatur diskutiert und in unterschiedlichen Anwendungsbereichen verwendet wurden. Sein Ziel ist es herauszustellen, welche der entwickelten Ansätze geeignet sind, den 'Herausforderungen des 21. Jahrhunderts' gerecht zu werden (vgl. Henry 2000: 1; Stufflebeam 2001: 8).

Hierzu ordnet Stufflebeam die von ihm untersuchten Ansätze vier Bereichen zu: pseudo-Evaluationsansätze, die fehlerhafte oder unvollständige Ergebnisse erzeugen und nicht weiter in die Betrachtung einfließen sowie 'Questions and/or Methods Oriented, Improvement/Accountability and Social Agenda/Advocacy Approaches'.

▪ Fragen- und methodenorientierte Evaluation

Alle Ansätze, die diesem Bereich zugeordnet werden, bezeichnet Stufflebeam auch als 'Quasi-Evaluationsstudien', da sie nur selten Belege liefern, die die Leistung oder den Wert eines Programms vollständig beurteilen lassen; in der Mehrzahl der Fälle ist ihr Fokus zu eng oder die Ausrichtung nur bedingt geeignet, die Programmleistungen vollständig zu beurteilen.

Während fragenorientierte Programmevaluationsansätze einen eng definierten Bereich von Fragestellungen, spezifiziert aus Zielen des Programms, mit einem breit angelegten Methodenrepertoire zu beantworten suchen, verwenden methodenori-

enterte Ansätze in der Regel nur eine Methode; Ausgangsposition kann hierzu der Entwurf eines kontrollierten Versuchs, ein genormter Test oder die Theorie eines Programms (o. a.) sein. Die Programmtheorie basierten Evaluationsansätze ordnet Stufflebeam der Gruppe der methodenorientierten Ansätze zu, da er das 'Theoriekonstrukt' eines Programmes/Modellversuch als methodische Variante ansieht, die die Evaluation anleitet (vgl. Stufflebeam 2001: 16f).

▪ **Verbesserungs- und verantwortungsorientierte Evaluation**

Die Ansätze dieses Bereiches betonen gleichsam die Notwendigkeit der vollständigen Beurteilung der Stärken und Schwächen eines Programms. Ausgehend von den Bedürfnissen der Stakeholder, die die Kriteriengrundlage für die Evaluation des Programms bilden, ziehen sie alle relevanten (technischen und ökonomischen) Aspekte heran, um Programmpläne und -verläufe zu prüfen. Dabei erfolgt keine Beschränkung auf diejenigen Erkenntnisse, die in engem Zusammenhang mit den Programmzielen stehen, sondern es werden auch nicht-intendierte Entwicklungen und Ergebnisse untersucht. Auf der Suche nach unzweifelhaften Erkenntnissen kommen verschiedene qualitative und quantitative Methoden zum Einsatz. Die Leistung dieser Ansätze liege in der Generierung von Entscheidungshilfen bei der Beurteilung von zur Auswahl stehenden Programmen, indem sie den Stakeholdern helfen, die Leistungen der Programme zu prüfen (vgl. Stufflebeam 2001: 42).

▪ **Sozialverträglichkeits- und Stellvertreter-Ansätze**

Ziel dieser Ansätze ist es, Chancengleichheit in Bezug auf Bildungsaspekte herzustellen. In ihrer konstruktivistischen Ausrichtung greifen sie die Perspektiven sowohl der Stakeholder als auch der von Bildungsexperten für die Untersuchung und die Beurteilung von Programmen auf; zum Einsatz kommen dabei vor allem qualitative Methoden. Als Ergebnis wird keine 'richtige' oder 'beste' Lösung angestrebt, vielmehr reflektieren die Ansätze den 'kulturellen Pluralismus, die moralische Relativität' sowie die Möglichkeit 'multipler Realitäten' (vgl. Stufflebeam 2001: 62f).

Jeder der Ansätze in den vier Bereichen wird durch zehn charakterisierende Merkmale beschrieben:

Ursache/Anlässe, die Evaluatoren dazu führen, eine Evaluationsstudie zu initiieren; Hauptanwendungsbereich des Ansatzes; Quellen der Untersuchungsfragen; Fragen, die charakteristisch für den jeweiligen Untersuchungstyp sind; Methoden; Personen, die den jeweiligen Ansatz konzeptionalisiert haben; weitere Personen, die zur Verbreitung des Ansatzes beigetragen haben; Schlüsselüberlegungen, die bei der Wahl des einzusetzenden Ansatzes anzustellen sind; Stärken und Schwächen.

Auf Grundlage einer tabellarischen Gegenüberstellung (Stärken-, Schwächenanalyse) werden Schlussfolgerungen gezogen, die deutlich machen, welcher Ansatz für die weitere Verwendung ungeeignet bzw. welcher besonders leistungstark ist. Als Ergebnis verbleiben neun Ansätze, die Stufflebeam besonders geeignet für eine weitere Entwicklung/Verwendung scheinen, so dass diese, auf Basis der vom 'Joint Committee on Standards for Educational Evaluation' entworfenen Evaluationsstandards, einer weiteren Meta-Evaluation unterzogen werden. Zwecks Opera-

tionalisierung werden hierfür die dreißig Standards des Joint Committee durch zehn 'Schlüsselmerkmale' untersetzt³¹ und auf jeden der neun Ansätze angewandt. Ungeachtet der potenziellen Gesamteinschätzung des einzelnen Evaluationsansatzes werden diese als vollständig ungeeignet beurteilt, wenn sie im Hinblick auf einen der von Stufflebeam als 'unerlässlich' eingeschätzten Standards ('K1 – Unterstützung der Dienstleistungsorientierung', 'G5 – Valide Informationen', 'G10 – Begründete Schlussfolgerungen' und 'G11 – Unparteiische Berichterstattung') als 'schwach' eingeschätzt werden.

4.4.2 Stufflebeams Einschätzung programmtheorieorientierter Evaluationsansätze

Der Erfolg eines theoriegeleiteten Ansatzes, und damit auch die Aussagefähigkeit der Evaluation, hängen im Wesentlichen von der Qualität des zu Grunde liegenden Theoriekonstrukts ab. Notwendig sei die Konzeptualisierung einer kontextabhängigen Theorie sowie die Formulierung und Überprüfung von aus der Theorie abgeleiteten Hypothesen. Jedoch hielten es Evaluatoren in vielen Fällen nicht für möglich, vollständig die Bandbreite der 'Theorieentwicklungs- und Gültigkeitsstufen' im Rahmen der Zeit- und Personalressourcen zu evaluieren, da der Aufwand sehr hoch sei:

"The more demanding grounded theory requires a systematic, empirical process of observing event or analyzing materials drawn from operating programs, followed by an extensive modeling process (Stufflebeam 2001: 38)".

Basis für eine theoriegeleitete Programmevaluation sei häufig ein Modell der 'Programmlogik'. Hierbei könne es sich um einen detaillierten 'Flowchart' handeln, der Aussagen über die vermutete Verarbeitung von 'Programm-Inputs' gibt, die Einfluss auf die beabsichtigten Outcomes haben oder um eine 'Netzwerkanalyse' auf Grundlage der 'Grounded Theory'. Letztere stellt eine Auswertungstechnik zur Entwicklung und Überprüfung von Theorien dar, die sich auf vorhandenes Material stützt und in den Daten 'verankert' ist. Ziel ist das Aufzeigen der Vernetzung von Kategorien und Subkategorien innerhalb der untersuchten Theorie (vgl. Bortz 1995: 308).

Für die Evaluation sei es sinnvoll, das Ausmaß der theoretischen Fundierung der Programmpläne und Arbeitsweisen bzw. die Ausprägung eines Modells zu untersuchen. Sollte eine angemessene Theorie vorhanden sein, könne der Evaluator diese nutzen, um die Evaluation zu strukturieren und durchzuführen. Jedoch rät Stufflebeam davon ab, dass Evaluatoren selbst eine Theorie oder Programmlogik entwickeln, wenn diese nicht existiert. Das Ergebnis könnten Ressourcenprobleme sein oder die Generierung einer nicht 'angemessenen' Theorie, d. h. entweder eines Konstrukts, das nicht mit dem realen Programm übereinstimmt oder die Anforderungen an eine 'gute' Theorie vernachlässigt. Dennoch sei es sinnvoll, in einem begrenzten zeitlichen Rahmen, ein Programm-Modell zu formen, um Messvariablen zu identifizieren. Voraussetzung sei es jedoch, dieses Modell nicht als validierte Theorie anzusehen.

31) Stufflebeam unterscheidet vier Ausprägungsgrade je Standard: schwach, angemessen, gut, sehr gut und außergewöhnlich gut/exzellente.

Trotz seiner generell positiven Einschätzung theoriebasierender Evaluationen kommt Stufflebeam zu der Erkenntnis, dass es im Hinblick auf die Anwendungspraxis keine fundierten Gründe gäbe, eine solche Evaluation zu empfehlen, da diese wegen der häufig nicht vorhandener Theoriebasis nicht durchführbar bzw. die Gefahr des Fehlschlagens wegen bestehender Ressourcenbeschränkungen hoch sei (vgl. Stufflebeam 2001: 38f).

4.5 Zweite Zwischenbilanz: Der Stand der Forschung in der Evaluationsliteratur

Hinsichtlich des Versuchs der definitorischen Fassung des Begriffes 'Evaluation' werden in der Vielzahl der Begriffsbeschreibungen vor allem zwei Komponenten in der Literatur hervorgehoben, die den Begriffgegenstand charakterisieren – Informationssammlung über das Evaluationsobjekt (Daten-Generierung) und die Beurteilung des Evaluationsgegenstandes auf Grundlage von Wertestandards.

Darüber hinaus sind verschiedene Merkmalskataloge vorhanden, die aus variierenden Deskriptions-Merkmalen von Evaluation als Entscheidungs- oder Strategie-Instrument einerseits und/oder aus Merkmalen der zu evaluierenden Maßnahmen andererseits bestehen (vgl. Westermann 2002: 5ff).

Da die genannten Vorgehensweisen bei der Beschreibung des Begriffes Evaluation wenig zielführend, weil inhaltlich beliebig veränderbar sind, ist alternativ dazu (als Diskussionsvorschlag für die Evaluationsforschung in Kap. 4.1) ein 'dimensionaler Ansatz' gewählt worden, in dem die in der Literatur als konsensuale Kernkomponenten beschriebenen Evaluationsmerkmale (vgl. Wottawa, Thierau 1998: 14) als Ausprägungen von differenzierbaren 'Variablen' angesehen und im Hinblick auf das jeweilige Evaluationsvorhaben expliziert werden (müssen). Im Ergebnis entstehen hierdurch zahlreiche kontextbezogene 'Evaluationsprofile', auf Basis eines modularen Evaluations-Begriffskomplexes, deren Ausgestaltung die Bandbreite der Entscheidungen bei der Konzeptionierung eines Evaluationsvorhabens verdeutlichen. Dieses Vorgehen charakterisiert **'Evaluation' als ein mehrdimensionales kontextbezogenes Instrument**, das nicht als standardisiertes und reproduzierbares Vorhaben angesehen werden darf und sich damit einer generalisierenden Definition oder endlichen Merkmalsaufzählung entzieht.

Der von Clarke, Suchmann u. a. verwendete Begriff der Evaluationsforschung als Bezeichnung für (wissenschaftlich gestützte) Evaluation suggeriert (fälschlicherweise s. u.), dass es sich hierbei um Forschung über Evaluation handle (vgl. Clarke 1999; Suchmann 1967); zum Teil wird in diesem Zusammenhang auch von 'angewandter Forschung' gesprochen. Der Vergleich der Implikationen von **Forschung und Evaluation** (siehe Kap. 4.1) hat jedoch die jeweils **unterschiedliche Ausrichtung** hervorgehoben, so dass die Konzentration auf Evaluationsvorhaben mit dem Ziel, auch einen Beitrag zur Evaluationsforschung leisten zu können, zunächst nur auf ein singuläres Phänomen bezogenes, interpretiertes und bewertetes Datenmaterial generiert. Die Verwendung solcher 'Erkenntnisse' kann jedoch wegen mangelnder Objektivität und fehlender Wertfreiheit nicht oder nur bedingt zu einer elaborierten Evaluationsforschung als wissenschaftlicher Disziplin beitragen.

In der amerikanischen Evaluationsliteratur wird als der forschungsbezogener Tätigkeitsschwerpunkt die Beschäftigung mit der methodischen Vorgehensweise bei Evaluationen hervorgehoben: "Dabei können zwei grundsätzlich verschiedene Orientierungen unterschieden werden [...] (vgl. Chen 1990; Clarke 1999). Die eine wird als quantitative (experimentelle oder naturwissenschaftliche) Orientierung beschrieben, die andere als qualitative (naturalistische, konstruktivistische, interpretatorische, holistische oder alternative) Orientierung" (Westermann 2002: 23). Westermann stellt in diesem Zusammenhang fest, dass diese methodischen Orientierungen festen, nicht hinterfragten Paradigmen im Sinne Kuhns entsprechen (siehe hierzu auch Kuhn 1981).

Zur Unterscheidung dieser Paradigmen kann u. a. deren methodologische Orientierung hervorgehoben werden:

Während quantitative Ansätze durch eine 'deduktive Methodologie' gekennzeichnet und auf die Falsifizierung aus Theorien abgeleiteter Hypothesen mittels empirischer Daten abzielen, basieren qualitative Ansätze auf einer 'induktiver Methodologie', die auf einer zu Grunde gelegten breiten Informationsbasis versucht, allgemeine Zusammenhänge und Regeln abzuleiten (vgl. Westermann 2002: 23). Ausprägungen und Entwicklungsstufen hierzu spiegeln sich in chronologischen und typologischen Darstellungen wider (siehe Kap. 4.2).

Die aus der Methodologiediskussion entstandene starke Methodenorientierung in der Evaluationsforschung hat zum Vorwurf der einseitigen wissenschaftlichen Orientierung geführt, so dass Evaluationen als Vorhaben bezeichnet werden, denen eine theoretische Fundierung im Sinne von Annahmen über Aktivitäten, Prozessen, Zielen und Beziehungen zwischen den Programm-/Modellversuchsaktivitäten fehlen. Das Ergebnis sind Evaluationen, die nicht auf erklärende oder vermittelnde Variablen aufbauen und somit als Black-Box-Modelle der Evaluation bezeichnet werden und keine Erklärung bieten können, wie ein Programm, ein Modellversuch funktioniert, d. h. wie die intendierten und nicht-intendierten Effekte entstehen.

Eine Möglichkeit der Ermittlung solcher Effekte/Wirkungen über den Ablauf von Modellversuchen bietet das Modell der 'theoriebasierten Evaluation'. Der Begriff ist dahin gehend irreführend, als mit 'Theorie' kein System abstrakter und generalisierender Hypothesen mit räumlich und zeitlich uneingeschränktem Geltungsanspruch gemeint ist (vgl. Kromrey 2001: 123), sondern vielmehr eine 'Theorie des Programmablaufs' wie im 'grounded-theory-Konzept' (siehe Weiss 1995, 1997). Kromrey schlägt daher alternativ, in Anlehnung an Patton, den Terminus 'logisches Modell' vor.

Methodische Voraussetzung für eine theoriebasierte Evaluation ist ein System von operationalisierten Hypothesen, das die Planung, Implementation und Messung der Effekte rekonstruieren lässt. Chen hebt in seinem Modell in Bezug auf die Konstruktion einer Programmtheorie die Spezifikation der entsprechenden Variab-

len auch hervor (siehe Kap. 4.3.2). Doch auch die entwickelten Ansätze zur Programmtheorie, wie die von Chen 1990, Birckmayer und Weiss 2000, Patton 1997 u. a. können nicht zur Qualitätssteigerung von Evaluationen beitragen, wenn Programme/Modellversuche dennoch keine (explizite) theoretische Fundierung aufweisen, d. h. nicht auf die vorhandenen wissenschaftlichen Erkenntnisse zurück gegriffen wird; Stufflebeam kommt im Rahmen der Meta-Evaluation etablierter Evaluationsansätze, bezugnehmend auf theoriebasierte Evaluationen, zu eben dieser Einschätzung (vgl. Stufflebeam 2001: 38).

Zu einem ähnlichen Ergebnis kommt auch Kromrey, der davon ausgeht, dass eine einheitliche Programmtheorie auf Basis einer rationalen Ziel- und Maßnahmenplanung nicht existiert. Vielmehr sieht er in einem 'Rationalmodell der theoriebasierten Programmevaluation' das Konstrukt der Evaluateure, das diese an das Programm/Modellversuch herantragen, um das Evaluationsdesign entwickeln zu können. Ungeachtet der Ressourcenproblematik, die Stufflebeam betont, geht Kromrey zudem von mehreren existierenden Programmtheorien für jedes Programm/Modellversuch aus, da sowohl die Evaluateure als auch die Projektleiter und die Modellversuchsakteure eigene Vorstellungen entwickeln, wie Modellversuchsmaßnahmen im Detail, unter den jeweils gegebenen Randbedingungen, zu organisieren sind. Nur im Idealfall kann vermutlich von übereinstimmenden resp. ergänzenden Programmtheorien ausgegangen werden, die den Fortgang des Modellversuchs steuern. Entscheidend ist es, diese im Regelfall divergierenden Programmtheorien zu rekonstruieren und in einem 'handlungslogischen Rahmenmodell' zu systematisieren (vgl. Kromrey 2001: 124). Nur so besteht die Möglichkeit, den gesamten Modellversuchsprozess aus Sicht der verschiedenen Akteure zu evaluieren und das Wirkungsfeld des Modellversuches nicht als 'black-box' zu belassen.

Vor dem Hintergrund dieser Einschätzungen und auf Basis der vorhandenen theoretisch fundierten Evaluationsansätze, könnte demnach zwar von einer (elaborierten) Evaluationsforschung gesprochen werden, die die geforderten theoretisch fundierten Ansätze aufzeigt und eine ausgereifte Methodologie vorhält; jedoch kann auch festgestellt werden, dass diese Überlegungen nach Ansicht von Stufflebeam und Kromrey nicht ausreichend verwendet werden, so dass nach wie vor von einer im Wesentlichen atheoretischen Evaluationspraxis, d. h. eine Evaluationspraxis, die nicht auf Basis einer rationalen Ziel- und Maßnahmenplanung (s. o.) agiert, ausgegangen werden muss.

5 Entwicklung des Evaluationskonzeptes

5.1 Konzeptspezifikation und Operationalisierung

Modellversuche stellen komplexe Projekte dar, die das Ziel verfolgen, Ausschnitte der sozialen Realität, wie die des beruflichen Bildungswesens (z. B. Curricula, methodische Orientierungen u. a.), zu verändern. Diese sind in Anlehnung an Widmer als Konstruktionen³² erster Ordnung zu beschreiben, die durch Konstruktionen zweiter Ordnung, durch Evaluationen, untersucht und bewertet werden (vgl. Widmer 1995: 9).

Da jedoch die Evaluationsforschung durch die beschriebenen (kritikbehafteten) heterogenen methodologischen Orientierungen charakterisiert ist (vgl. Kap. 4) und diese darüber hinaus Implikationen aus verschiedenen wissenschaftlichen Disziplinen enthalten, entsteht das Problem einer über verschiedene Paradigmen und Disziplinen hinweg anwendbaren und anerkannten Kriterienauswahl, die als ausgewogenes Konglomerat von Güte Merkmalen für die Metaevaluation herangezogen werden kann.

Als Hilfsmittel zur Untersuchung und Bewertung von Modellversuchen werden daher 'Evaluationsstandards' herangezogen, die als anerkannte Prinzipien zur Einschätzung der Qualität von Evaluationsvorhaben anzusehen sind (vgl. Sanders 1999: 24; Ridings, Stufflebeam 1981: 7; Widmer 1995: 10):

Ihre Funktion besteht einerseits in der Qualitätssicherung – hier gilt es durch den Meta-Evaluator zu prüfen, ob sich die wissenschaftliche Begleitung der Modellversuche hinsichtlich der Gestaltung ihres Evaluationsdesigns und der anschließenden –durchführung (Konstruktion zweiter Ordnung) an den Anforderungen der Standards orientiert hat, um somit die Qualität/Güte der Evaluation zu gewährleisten [Prüfstufe I; siehe Kap. 5.4]; andererseits übernehmen die Evaluationsstandards eine handlungsleitende Funktion, indem Operationalisierungen und Richtlinien (als Strategien zur Fehlervermeidung bei der Anwendung der Standards; vgl. Sanders 1999: 29) zu den Evaluationsstandards Hinweise zur Gestaltung von Evaluationen geben. Eine (derivative [Meta-] Funktion, d. h. eine abgeleitete, sich erst aus der Anwendung und Überprüfung der Evaluationsqualität ergebende) dritte Funktion liegt in der Rückwirkung der geprüften Evaluationsstandards auf die Metaevaluation (Konstruktion dritter Ordnung), da durch die Ausformulierung der Einschätzungen zu Konstruktionen zweiter Ordnung die Qualität der Metaevaluation selbst ersichtlich wird.

Das Joint Committee on Standards for Educational Evaluation hat vier Attribute/Standardgruppen definiert, die durch dreißig Einzel-Standards untersetzt werden und somit dazu beitragen, die vier Attribute zu definieren:

32) Der Begriff der 'Konstruktion erster, zweiter und dritter Ordnung' verdeutlicht, dass es sich hierbei um geplante und auf theoretischen Überlegungen basierende Planungs-, Durchführungs- oder Evaluationskonzepte handelt, die einen mittelbaren oder unmittelbaren Bezug untereinander aufweisen und den gleichen Ausschnitt der sozialen Realität (Gegenstand des Modellversuchs) reflektieren.

- **Nützlichkeitsstandards –**

zielen darauf ab, die Informationsbedürfnisse der Beteiligten und Betroffenen (B&B; Stakeholder) zu ermitteln und die gewonnenen Informationen verständlich und zeitnah den B&B mitzuteilen.

- **Durchführbarkeitsstandards -**

sollen die Praktikabilität und eine angemessene Ressourcenverwendung sicherstellen.

- **Korrekttheitsstandards -**

heben die ethische und rechtliche Orientierung bei der Gestaltung und Durchführung einer Evaluation hervor.

- **Genauigkeitsstandards -**

beziehen sich erstens auf eine umfassende Berücksichtigung 'erkennbarer' Modellversuchsbestandteile, zweitens auf die Erhebung derjenigen Daten, die geeignet sind, die Güte und Verwendbarkeit des Modellversuchs beurteilen zu können und drittens auf die Orientierung an Gütemerkmalen wissenschaftlichen Arbeitens (vgl. Sanders 1999: 28f).

Jedoch können die vom Joint Committee on Standards for Educational Evaluation aufgestellten, dimensionierten und hier präferierten Standards nicht in ihrer Gesamtheit für jeden Zweck genutzt werden, sondern die Sachdienlichkeit des jeweiligen Verwendungskontextes ist für die Auswahl entscheidend (vgl. Sanders 1999: 24.). Bei der Festlegung der zu verwendenden Standards können im Rahmen einer Metaevaluation nur solche zum Einsatz kommen, die anhand der schriftlichen Evaluationsdokumente (Zwischen- und Endberichte, evtl. auch Modellversuchsanträge) prüfbar sind. Um dem Grundsatz 'steigende Anzahl anwendbarer Evaluationsstandards führt zu steigender Qualität der Evaluation' zu folgen, wird nicht der Einsatz, sondern vielmehr der Ausschluss einzelner, vom Joint Committee on Standards for Educational Evaluation festgelegter Standards mit Hilfe einer Negativ-Legitimation zu begründen sein (siehe Kap. 5.2).

Nach der Auswahl werden die zum Einsatz kommenden Evaluationsstandards in einem ersten Schritt ohne spezifische Wertabstufung innerhalb der Standards und einem zweiten Schritt durch Untersetzung mit Ausprägungsgraden operationalisiert (siehe Anhang zu Teil B), so dass die Anwendung eines formalisierten Bewertungs- und Entscheidungsverfahrens möglich wird, da die einzelnen Standardausprägungen Nutzenwerten entsprechen. Bei der Einstufung von Modellversuchen hinsichtlich der Ausprägung der ihnen immanenten Sachverhalte im Rahmen generalisierter Dimensionen (Ausprägungsgrade) wird in dieser Arbeit ein ordinalskaliertes Niveau erreicht; demnach besteht die Möglichkeit der Rangordnung der Modellversuche in Bezug auf die untersuchten Dimensionen (Standardbereiche) (siehe hierzu u. a. Schnell u. a. 1999: 137; Bortz, Döring 1995: 67; Kromrey 2000: 209f).³³

33) Im Hinblick auf die Ausführungen zur Evaluationsdiskussion (Kap. 4/4.3), insbesondere zur Evaluation programmtheoretischer Ansätze als 'Status quo' der Erkenntnisse in diesem

5.2 Bestimmung der anzuwendenden Evaluationsstandards per Ausschlussverfahren

Die im Metaevaluationsprozess nicht zum Tragen kommenden Standards lassen sich wie nachfolgend kategorisieren, wobei eine trennscharfe Unterscheidung nicht immer möglich ist. Die Einordnung der Standards erfolgt dort, wo ein Schwerpunkt ihrer Wirksamkeit zu sehen ist³⁴:

▪ Die Standards enthalten überwiegend personenabhängige Aspekte

N2	Glaubwürdigkeit der Evaluatorsin: Wer Evaluationen durchführt, sollte sowohl vertrauenswürdig als auch kompetent sein, damit bei den Evaluationsergebnissen ein Höchstmaß an Glaubwürdigkeit und Akzeptanz erreicht wird.
	Ausschlussbegründung: Inwiefern ein Evaluationsmitarbeiter glaubwürdig auf die Gruppe der Betroffenen und Beteiligten wirkt, hängt im Wesentlichen von seinem persönlichen Auftreten ab, das im Evaluationsprozess zum Tragen kommt. Seine fachliche Kompetenz entzieht sich darüber hinaus häufig der Beurteilung.

▪ Evaluationsprozessbezogene Aspekte entziehen sich grundsätzlich oder zumindest der nachträglichen Betrachtung

N6	Rechtzeitigkeit und Verbreitung des Berichts: Wichtige Zwischenergebnisse und Schlussberichte sollten den vorgesehenen Nutzern so zur Kenntnis gebracht werden, dass diese sie rechtzeitig verwenden können.
	Ausschlussbegründung: Der Termin der Berichtslegung wird durch die Administration, hier der BLK, festgelegt und kann damit von subjektiven Notwendigkeitseinschätzungen der Nutzer abweichen.

mehrere Wissenschaftsbereiche tangierenden Themengebiet, sowie der Einschätzung Stufflebeams hierzu (Programmtheorie als strukturgebende Evaluationsgrundlage) wäre auch eine alternative Herangehensweise in einer metaevaluativen Untersuchung ausgewählter Modellversuche möglich, um die These einer 'atheoretischen Evaluationspraxis' (siehe Kap. 4) zu verifizieren. Hierzu müsste von folgender Hypothese ausgegangen werden: Konstruktionen zweiter Ordnung (Evaluationen) explizieren die in Konstruktionen erster Ordnung (Modellversuche) zumindest implizit enthaltene theoretische Fundierung als Grundlage für eine wirkungsbezogene Evaluation (vgl. Chen 1990: 66f). Eine Metaevaluation (Konstruktion dritter Ordnung) ist in diesem Zusammenhang als Evaluation der programmtheoretischen Basis von Konstruktionen zweiter Ordnung (Evaluation der Modellversuche) zu verstehen. Der Untersuchungsansatz wäre hierzu derart zu gestalten, dass die zu Grunde liegende theoretische Basis inklusive der einflussnehmenden Hypothesen expliziert und geprüft würde. Das Vorgehen in dieser Arbeit bietet hierzu lediglich Indizien, Hinweise also, wie sie z. B. durch den Standard K5 gegeben werden.

34) Bei den Bezeichnungskürzeln (z. B. N2) handelt es sich um die Kürzel des Joint Committee.

N7**Wirkung der Evaluation:**

Evaluationen sollten so geplant, durchgeführt und dargestellt werden, dass die B & B³⁵ dazu ermuntert werden, dem Evaluationsprozess zu folgen, damit die Wahrscheinlichkeit steigt, dass die Evaluation genutzt wird.

Ausschlussbegründung:

Motivationale Aspekte lassen sich retrospektiv kaum zufrieden stellend eruieren.

K3**Schutz individueller Menschenrechte:**

Evaluationen sollten so geplant und durchgeführt werden, dass die Rechte und das Wohlergehen der Menschen respektiert und geschützt werden.

K4**Human gestaltete Interaktion:**

Evaluatorinnen sollten in ihren Kontakten mit Anderen die Würde und den Wert der Menschen respektieren, damit diese nicht gefährdet und geschädigt werden.

Ausschlussbegründung:

Die Berücksichtigung der Rechte und des Wohlergehens der Beteiligten und Betroffenen sowie die Anerkennung der Würde kann, wenn überhaupt, nur mit hohem Aufwand nachträglich festgestellt werden, vor allem, wenn es sich hierbei um ehemalige Schüler handelt. Ausnahme: es handelt sich um eklatante Rechtsverstöße, die bekannt sind.

G7**Systematische Informationsprüfung:**

Die in einer Evaluation gesammelten, aufbereiteten und präsentierten Informationen sollten systematisch überprüft und alle gefundenen Fehler sollten korrigiert werden.

Ausschlussbegründung:

Inwiefern Informationen systematisch überprüft wurden, ist aus der heutigen Perspektive nicht angemessen prüfbar.

G11**Unparteiische Berichterstattung:**

Die Verfahren der Berichterstattung sollten mittels Vorkehrungen gegen Verzerrungen durch persönliche Gefühle und Vorlieben irgendeiner Evaluationspartei geschützt werden, so dass Evaluationsberichte die Ergebnisse fair wiedergeben.

Ausschlussbegründung:

Inwieweit Darstellungen durch persönliche Präferenzen beeinflusst sind, lässt sich zumeist nur in besonders extremen Fällen aufzeigen.

35) B&B=Betroffene und Beteiligte

▪ **Berichte dokumentieren geforderte Aspekte i. d. R. nicht**

D2

Politische Tragfähigkeit:

Evaluationen sollten mit Voraussicht auf die unterschiedlichen Positionen der verschiedenen Interessengruppen geplant und durchgeführt werden, um deren Kooperation zu erreichen und um mögliche Versuche irgendeiner dieser Gruppen zu vermeiden, die Evaluationsaktivitäten einzuschränken oder die Ergebnisse zu verzerren respektive zu missbrauchen.

Ausschlussbegründung:

Selbstverständlich müssen die Interessen aller Beteiligten angehört und, wenn möglich, berücksichtigt werden, ob dies allerdings mit einer Voraussicht geschieht, die eine Kooperation aller mit sich bringt, kann auf der Basis der Berichterstattung nicht mit hinreichender Sicherheit geprüft werden. Lediglich aus den Äußerungen über Hindernisse und Schwierigkeiten können Rückschlüsse gezogen werden. Jedoch kann auf diesem Wege nicht auf die Einhaltung dieses Standards geschlossen werden.

D3

Kostenwirksamkeit:

Die Evaluation sollte effizient sein und Informationen mit einem Wert hervorbringen, der die eingesetzten Mittel rechtfertigt.

Ausschlussbegründung:

Die Kostenwirksamkeit im Sinne einer Effizienz von eingesetzten Mitteln und Informationsgewinnung lässt sich nur mit einem hohem Aufwand und darüber hinaus unter hoher Unsicherheit ermitteln.

K2

Formale Vereinbarungen:

Die Pflichten der Vertragsparteien einer Evaluation [...] sollten schriftlich festgehalten werden, damit die Parteien verpflichtet sind, alle Bedingungen dieser Vereinbarung zu erfüllen oder aber diese erneut zum Gegenstand von formalen Vereinbarungen zu machen.

Ausschlussbegründung:

Formale Vereinbarungen werden nicht (zumindest nicht systematisch) in die Berichterstattung einbezogen.

K8

Finanzielle Verantwortlichkeit:

Die Zuweisung und Ausgabe von Ressourcen durch die Evaluatorin sollte durch eine sorgfältige Rechnungsführung nachgewiesen werden und auch anderweitig klug sowie ethisch verantwortlich erfolgen, damit die Ausgaben verantwortungsbewusst und angemessen sind.

Ausschlussbegründung:

Auch hier fehlen Informationen in der Berichterstattung, die auf eine angemessene und verantwortungsbewusste Verwendung schließen lassen.

Im Folgenden (siehe Kap. 5.3) werden die ausgewählten Evaluationsstandards inklusive der vorgenommenen Operationalisierungen zunächst ohne spezifische Wertabstufungen innerhalb der Standards dargestellt. Aus Gründen der Übersichtlichkeit werden die Ausprägungsgrade nicht mit angegeben. Eine vollständige Übersicht findet sich im Anhang zu Teil B.

5.3 Ausgewählte Evaluationsstandards zur Anwendung im Rahmen der metaevaluativen Dokumentenanalyse

N1

Ermittlung der Beteiligten und Betroffenen:

Die an einer Evaluation beteiligten oder von ihr betroffenen Personen sollten identifiziert werden, damit deren Interessen und Bedürfnisse berücksichtigt werden.

Operationalisierung:

Die B&B werden, sofern es sich um potenziell wichtige Personen handelt und die zeitlichen und finanziellen Beschränkungen dies zulassen, identifiziert, ihre Informationsbedürfnisse ermittelt und in die Planung und Durchführung mit einbezogen. Weniger einflussreichen Gruppen wird eine ausreichende Beteiligung ermöglicht.

N3

Umgang und Auswahl der Informationen:

Die gewonnenen Informationen sollten von einem Umfang und einer Auswahl sein, welche die Behandlung sachdienlicher Fragen zum Modellversuch ermöglichen und gleichzeitig auf die Interessen und Bedürfnisse des Auftraggebers und anderer B&B eingehen.

Operationalisierung:

Die Evaluatoren bestimmen das Informationspotenzial und selektieren diejenigen Informationen, die im Sinne des Modellversuchs wichtig sind und sondern sie somit aus der Informationsflut heraus.

Darüber hinaus werden auch Informationen ausgewählt, deren Bedeutung von den Auftraggebern bislang übersehen wurde.

N4

Feststellung von Werten:

Die Perspektiven, Verfahren und Gedankengänge, auf denen die Interpretationen der Ergebnisse beruhen, sollten sorgfältig beschrieben werden, damit die Grundlage der Werturteile klar ersichtlich ist.

Operationalisierung:

Die der Interpretation zugrunde gelegten Wertmaßstäbe werden offengelegt und die Perspektiven, Verfahren, Gedankengänge werden beschreibend dokumentiert, bevor auf deren Grundlage Deutungen vorgenommen und Urteile ausgesprochen werden.

N5**Klarheit des Berichts:**

Evaluationsberichte sollten den evaluierten Modellversuch einschließlich seines Kontextes ebenso beschreiben wie die Ziele, die Verfahren und Befunde der Evaluation, damit die wesentlichen Informationen zur Verfügung stehen und leicht verstanden werden können.

Operationalisierung:

Die B&B können leicht erkennen, welche Ziele die Evaluation verfolgte, was evaluiert wurde, wie die Evaluation durchgeführt wurde, welche Informationen gewonnen wurden, zu welchen Schlussfolgerungen man kam und ggf. welche Empfehlungen gemacht wurden.

N**Nützlichkeitsstandards**

Die Nützlichkeitsstandards sollen sicherstellen, dass sich eine Evaluation an den Informationsbedürfnissen der vorgesehenen Evaluationsnutzer ausrichtet.

Operationalisierung:

Die Evaluation ist derart aufgebaut, dass sie die Informationsbedürfnisse der B&B ermittelt und diese im Aufbau, der Durchführung und Auswertung auch berücksichtigt.

D1**Praktische Verfahren:**

Die Evaluationsverfahren sollten praktisch sein, so dass Störungen minimiert und die benötigten Informationen beschafft werden können.

Operationalisierung:

Die ausgewählten Evaluationsverfahren (Art der Ev.-instrumente, Einsatz und Informationssammlung) sind hinsichtlich ihres Aufwandes im Hinblick auf den zeitlichen Rahmen vertretbar.

K1**Unterstützung der Dienstleistungsorientierung:**

Die Evaluation sollte so geplant werden, dass Organisationen dabei unterstützt werden, die Interessen und Bedürfnisse des ganzen Zielgruppenspektrums zu berücksichtigen und ihre Tätigkeiten danach auszurichten.

Operationalisierung:

Die Auswirkungen eines Modellversuchs (Zielsetzungen) werden im Hinblick auf die Bedürfnisse der B&B untersucht. Daher wird vor allem auch in Zwischenberichten über Stärken und Schwächen des Modellversuchs und mögliche Verbesserungsvorschlägen berichtet.

K5**Vollständige und faire Einschätzung:**

Evaluationen sollten in der Überprüfung und in der Präsentation der Stärken und Schwächen des evaluierten Modellversuchs vollständig und fair sein, so dass die Stärken weiter ausgebaut und die Problemfelder angesprochen werden können.

Operationalisierung:

Es ist auf Stärken und Schwächen hinzuweisen, auch wenn diese unbeabsichtigt sind und jedes einzelne Ergebnis ist zu begründen. Darüber hinaus sind die Methoden der Evaluation in diesem Zusammenhang zu nennen, damit ihre Auswirkungen auf Ergebnisse und Schlussfolgerungen bewertet werden können.

K6**Offenlegung der Ergebnisse:**

Die Vertragsparteien einer Evaluation sollten sicherstellen, dass die Evaluationsergebnisse – einschließlich ihrer relevanten Beschränkungen – den durch die Evaluation betroffenen Personen ebenso wie alle jenen, die einen ausgewiesenen Anspruch auf die Evaluationsergebnisse haben, zugänglich gemacht werden.

Operationalisierung:

Sämtliche Handlungen, öffentliche Erklärungen und schriftliche Berichte sind den B&B und der interessierten Öffentlichkeit zugänglich zu machen. Es werden Ansprechpartner oder eine Zentralstelle genannt, bei der alle Informationen angefordert oder eingesehen werden können.

K7**Deklaration von Interessenkonflikten:**

Interessenkonflikte sollten offen und aufrichtig behandelt werden, damit sie die Evaluationsverfahren und –ergebnisse nicht beeinträchtigen.

Operationalisierung:

Interessenkonflikte werden im Hinblick auf ihre Ursachen dokumentiert und Maßnahmen, die vor Interessenkonflikten schützen, werden schriftlich zwischen den Betroffenen vereinbart.

K**Korrektheitsstandards:**

Die Korrektheitsstandards sollen sicherstellen, dass eine Evaluation rechtlich und ethisch korrekt durchgeführt wird und dem Wohlergehen der in die Evaluation einbezogenen und auch der durch die Ergebnisse betroffenen Personen gebührende Aufmerksamkeit widmet.

Operationalisierung:

Die Evaluation wird unter Beachtung rechtlicher Normen einerseits sowie gesellschaftlicher und persönlicher Werte und Normen andererseits durchgeführt. Es ist sichergestellt, dass dem Wohlergehen der B&B ausreichend Aufmerksamkeit geschenkt wird.

G1**MV-Dokumentation:**

Der zu evaluierende Modellversuch sollte klar und genau beschrieben und dokumentiert werden, so dass er eindeutig identifiziert werden kann.

Operationalisierung:

Aus der Dokumentation geht hervor, wie der MV ursprünglich geplant war und wie er tatsächlich umgesetzt wurde. Aus den Beschreibungen können Inhalte und Wirkungen miteinander in Beziehung gesetzt werden.

G2**Kontextanalyse:**

Der Kontext, in dem der Modellversuch angesiedelt ist, sollte ausreichend detailliert untersucht werden, damit mögliche Beeinflussungen des Modellversuchs identifiziert werden können.

Operationalisierung:

Der Begriff Kontext bezieht sich im Hinblick auf Planung, Durchführung und Bericht auf die Einordnung des MV in das politisch-soziale Umfeld (Kulturhoheit der Länder), den zeitlichen Rahmen, parallel laufende Aktivitäten, das Personal sowie weitere entscheidende Einflussfaktoren.

G3**Beschreibung von Zielen und Vorgehen:**

Die Zwecksetzungen und das Vorgehen der Evaluation sollten ausreichend genau dokumentiert und beschrieben werden, so dass sie identifiziert und eingeschätzt werden können.

Operationalisierung:

Die Ziele der Evaluation und die beabsichtigte Nutzung der Ergebnisse werden beschrieben und vor Beginn der Evaluation den Teilnehmern mitgeteilt. Nach Abschluss des MV erfolgt eine aktuelle Beschreibung, was mit welcher Absicht gemacht wurde, damit Ergebnisse und Empfehlungen beurteilt und interpretiert werden können.

G4**Verlässliche Informationsquellen:**

Die in einer Modellversuchsevaluation genutzten Informationsquellen sollten hinreichend genau beschrieben sein, damit die Angemessenheit der Informationen eingeschätzt werden kann.

Operationalisierung:

Die genutzten Informationsquellen werden beschrieben, die Verfahren der Informationsgewinnung werden dokumentiert und es wird deutlich, ob die Informationen einzigartig oder einseitig beeinflusst sind.

G5**Valide Informationen:**

Die Verfahren zur Informationsgewinnung sollten so gewählt oder entwickelt und dann umgesetzt werden, dass die Gültigkeit der gewonnenen Interpretationen für den gegebenen Zweck sichergestellt ist.

Operationalisierung:

Es ist sichergestellt, dass die Auswahl der Verfahren zur Informationsgewinnung begründet wird und diese derart gestaltet sind, dass die beabsichtigten Daten damit erhoben werden können.

G6**Reliable Informationen:**

Die Verfahren zur Informationsgewinnung sollten so gewählt oder entwickelt und dann umgesetzt werden, dass die Zuverlässigkeit der gewonnenen Informationen für den gegebenen Zweck sichergestellt ist.

Operationalisierung:

Die Zuverlässigkeit der EV-Verfahren wird dadurch sichergestellt, dass

bei der Auswahl der Verfahren die immanenten Zufallsfehler berücksichtigt, beschrieben und/oder reduziert werden.

G8**Analyse quantitativer Informationen:**

Quantitative Informationen einer Evaluation sollten angemessen und systematisch analysiert werden, damit die Fragestellungen der Evaluation beantwortet werden.

Operationalisierung:

Die zusammengetragenen quantitativen Daten sind nachvollziehbar geordnet, zusammengefasst, analysiert und interpretiert. Zum besseren Verständnis der Daten und der Darstellung und Interpretation der statistischen Ergebnisse sind außerdem graphische Darstellungen zu wählen. Insgesamt sind die Ergebnisse geeignet, den Einsatz der jeweiligen Evaluationsmethoden zu rechtfertigen.

G9**Analyse qualitativer Informationen:**

Qualitative Informationen einer Evaluation sollten angemessen und systematisch analysiert werden, damit die Fragestellungen der Evaluation beantwortet werden.

Operationalisierung:

Die gesammelten qualitativen Informationen sind durch mehr als eine Quelle belegt. Diese Informationen sind in Kategorien unterteilt und die Schlussfolgerungen und Empfehlungen beziehen sich auf diese Kategorien und sind geeignet, die Fragestellungen der Evaluation zu dokumentieren, zu erläutern und zu beantworten.

G10**Begründete Schlussfolgerungen:**

Die in einer Evaluation gezogenen Folgerungen sollten ausdrücklich begründet werden, damit die B&B diese einschätzen können.

Operationalisierung:

Die Schlussfolgerungen beantworten die Fragestellungen der Adressaten, spiegeln die Verfahren und Ergebnisse der Evaluation wider; sie werden begründet und gerechtfertigt. Auch andere Erklärungen der Ergebnisse werden erarbeitet, bewertet und dargelegt.

G12**Meta-Evaluation:**

Die Evaluation selbst sollte formativ und summativ in Bezug auf die vorliegenden oder andere wichtige Standards evaluiert werden, so dass die Durchführung entsprechend angeleitet werden kann und damit die Beteiligten und Betroffenen bei Abschluss deren Stärken und Schwächen gründlich überprüfen können.

Operationalisierung:

Parallel zum Evaluationsprozess des MV findet eine Meta-Evaluation (formativ und nach Abschluss summativ) statt, die durch externe Meta-Evaluatoren durchgeführt wird.

G**Genauigkeitsstandards:**

Die Genauigkeitsstandards sollen sicherstellen, dass eine Evaluation über die Güte und/oder die Verwendbarkeit des evaluierten Modellversuchs fachlich angemessene Informationen hervorbringt und vermittelt.

Operationalisierung:

Bei der Evaluation werden nur solche Methoden verwandt, die geeignet sind, die Güte und/oder die Verwendbarkeit des zu evaluierenden Modellversuchs zu beurteilen und auf diesem Wege für die Fragestellung relevante und ausreichende Informationen hervorbringt.

5.4 Kongruenzprüfung von evaluationsbezogenen Zielen und hierauf bezogene Evaluationsverfahren

Ungeachtet des Evaluands ist eine Funktion von Evaluationen (Konstruktionen zweiter und dritter Ordnung) die Wertzumessung eines Objektes/Modellversuches auf Basis von 'Diskrepanz-Informationen' zwischen ex-ante formulierten 'Soll-Zuständen' und ex-post festgestellten 'Ist-Zuständen' (vgl. Steinmetz 193: 80). Auch im vorliegenden Fall werden solche Diskrepanz-Informationen zwischen Evaluationsstandards und 'Güte/Leistung' der Evaluation erhoben (siehe Kap. 5.1), was zu einer Betrachtungsweise führt, die die Kongruenz zwischen den evaluationsbezogenen Zielen der wissenschaftlichen Begleitung und den hierauf bezogenen Evaluationsverfahren (zunächst) unberücksichtigt lässt und damit Aussagen über die Eignung der eingesetzten Verfahren unterbleiben muss. Pätzold weist jedoch darauf hin, "[...] dass die Beantwortung der Frage, was Modellversuche wirklich leisten, nicht nur von den Leistungskriterien, sondern auch von der Form der Leistungsfeststellung abhängt, d. h. der für die Erkenntnisgewinnung eingesetzten Erkenntnisverfahren" (Pätzold 1995: 57).

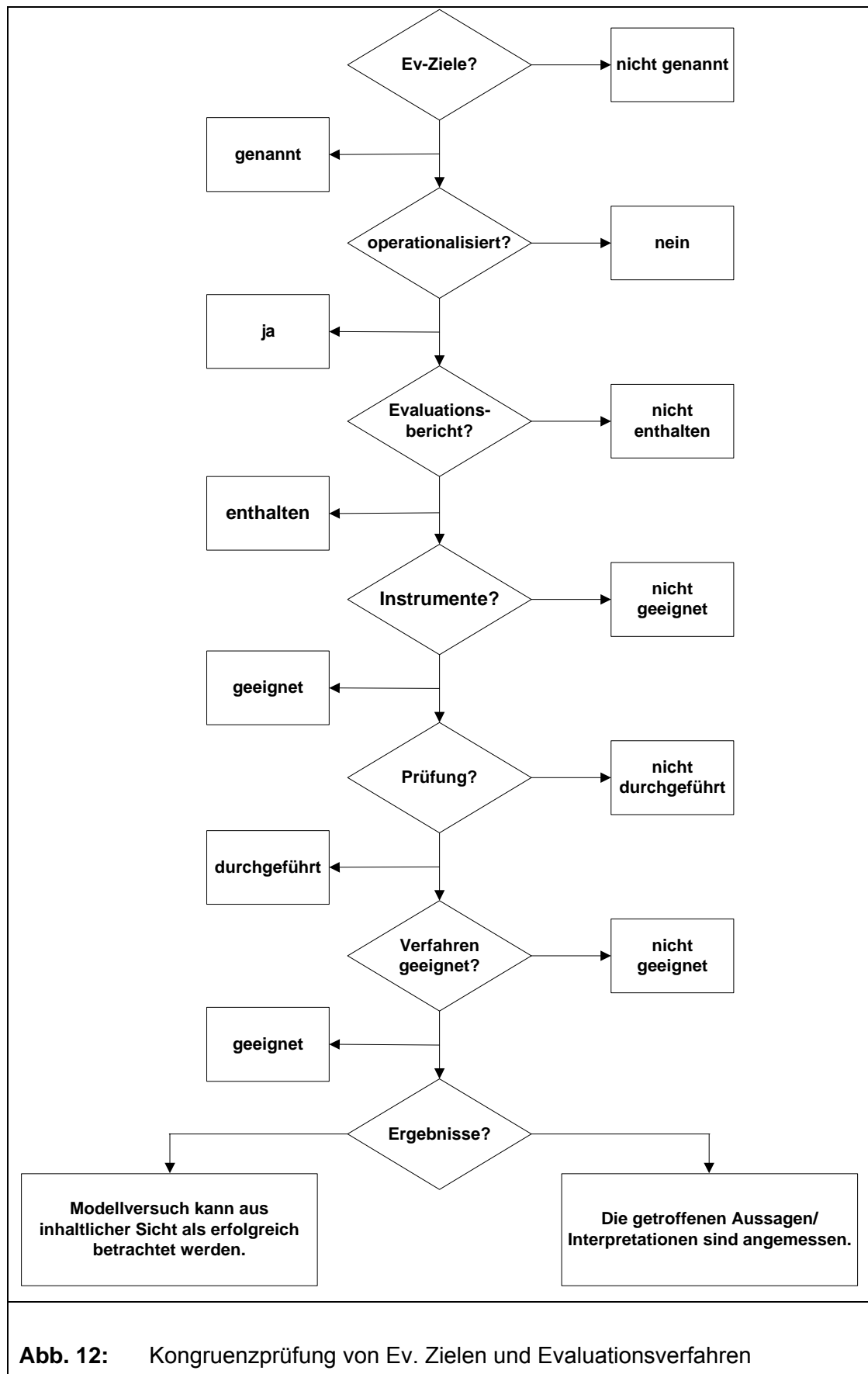
Dieses Zitat unterstreicht die Abhängigkeit von Aussagen über das Maß der Leistungserfüllung von Modellversuchen (Modellversuche werden an ihren transferfähigen Produkten gemessen ≈ Modellversuchsebene) einerseits von den Leistungskriterien, andererseits aber auch von den eingesetzten Evaluationsverfahren (≈Evaluationsebene). Sind letztere nicht oder nur bedingt geeignet, die 'Güte/Leistung' eines Modellversuches festzustellen, dann können keine gesicherten Erkenntnisse zur Leistung eines Modellversuchs generiert werden.

Daher ist die Evaluation von Konstruktionen zweiter Ordnung auf Basis von Evaluationsstandards durch ein Prüfschema zu ergänzen (siehe Abb. 12), das die Eignung der eingesetzten Verfahren zur Evaluation der Ziele der WBL (evaluationsbezogene Ziele) aufzeigt [Prüfstufe II], um hierauf aufbauend Angaben hinsichtlich der Angemessenheit der in den Modellversuchsdokumentationen getroffenen Aussagen/Interpretationen sowie der Einschätzung des Modellversuchserfolgs in Bezug auf die Gestaltung und Durchführung der Evaluation durch die wissenschaftlichen Begleitung vornehmen zu können. Sollte einer der sieben vorgesehenen Prüfschritte nicht durchgeführt werden können, da

Angaben in den Dokumentationen hierzu fehlen bzw. die eingesetzten Instrumente und Verfahren als ungeeignet angesehen werden müssen, dann ist wegen des fehlenden Nachweises der Kongruenz bzw. wegen mangelnder Angemessenheit von Instrumenten/Verfahren die Eignung des Evaluationsansatzes der wissenschaftlichen Begleitung und somit die Transferfähigkeit des Modellversuches in Frage gestellt. Zusammenfassend lässt sich das in Teil B der metaevaluativen Studie zur Anwendung kommende zweistufige Leistungsprüfverfahren wie folgt beschreiben: Ausgangspunkt ist eine metaevaluative Dokumentenanalyse auf Grundlage eines aus elaborierten Evaluationsstandards entwickelten ordinalskalierten Kategorienrasters (siehe Kap. 5.1; Prüfstufe I); die Funktionen der Qualitätssicherung und Handlungsleitung von Evaluationsstandards lassen die Metaevaluation die 'Aufbauorganisation' der Konstruktionen zweiter Ordnung prüfen. In einem zweiten Schritt erfolgt eine Kongruenzprüfung von evaluationsbezogenen Zielen der WBL und Evaluationsverfahren (Prüfstufe II); dieses Vorgehen kann im weiteren Sinne als Überprüfung der 'Ablauforganisation' bezeichnet werden, da neben der untersuchten Kongruenz offen gelegt wird, ob der Verlauf der Modellversuchsevaluation von der Zielformulierung bis zur Zielprüfung (Konstruktion zweiter Ordnung) nachvollziehbar und korrekt ist (siehe Abb. 12).

Der für diese Arbeit Anwendung findende modulare Evaluationsbegriff sowie die zusammenfassende Charakteristik des eigenen evaluativen Vorgehens lassen sich durch Subsumtion unter die obigen Ausführungen (Kap. 4.1 und 5.1) wie folgt beschreiben:

Das metaevaluative Vorhaben stellt eine systematische Untersuchung der Leistungen und des Wertes eines Modellversuches (Baustein A – Basisdefinition in Anlehnung an Mertens; Mertens 2001: 41) dar, bei dem Modellversuchsdokumentationen als schriftliche Evaluationsdokumente herangezogen werden (Baustein B – Gegenstandsdimension, Evaluand, 'Outputbezug'). Zur Verifizierung der These, dass (Meta-) Evaluationen von Konstruktionen zweiter Ordnung, Aussagen im Hinblick auf die Güte von Konstruktionen erster Ordnung zulassen (siehe Kap. 1.2), deren Auswirkung auf der Objektebene wirksam werden, kommen operationalisierte und mit Ausprägungsgraden untersetzte Evaluationsstandards zum Einsatz. Darüber hinaus prüft ein Kongruenzschema die Eignung der eingesetzten Evaluationsverfahren auf Ebene von Konstruktionen zweiter Ordnung im Hinblick auf die angestrebten evaluationsbezogenen Ziele (Baustein C – methodische Dimension). Bei der auf Evaluationsdokumenten basierenden Metaevaluation handelt es sich um eine summative Evaluation, deren Ergebnisse prospektiven Charakter inne haben (Baustein D – Raum-Zeit-Dimension).



6 Evaluation ausgewählter Modellversuche (Teil B der Studie)

6.1 Ausgewählte Modellversuche der Gruppe G4 – Innovationen in bestehenden Fächern oder Schulformen

Generelles Charakteristikum der ausgewählten Modellversuche ist der Aspekt der Veränderung/Innovation, der mit ihnen verbunden werden kann. Bei der näheren Betrachtung erscheint es sinnvoll, eine weitere Untergliederung vorzunehmen (siehe 3.1.2).³⁶ Die zur Untersuchung ausgewählten Erprobungsvorhaben (19 Stück) beziehen sich auf den Bereich 'Didaktik-Methodik', genauer auf den Aspekt der 'Lehr-Lernmethoden'. Ziel von Modellversuchen dieser Gruppe ist die Entwicklung didaktisch-methodischer Konzepte zur Förderung beruflicher Handlungskompetenz. Diese sollen sich einerseits auf die Einbeziehung von Informations- und Kommunikationstechnologien beziehen, darüber hinaus sind alternative Lehr-Lernorganisationsformen zu erproben und vorzustellen. Durchführungsorte sind jeweils Schulen in zehn ('alten') Bundesländer.³⁷

Um die erforderlichen Dokumentationen beziehen zu können, sind die Kultusministerien der betroffenen Bundesländer angeschrieben worden und um Übersendung aller erschienenen Zwischen- und Abschlussberichte gebeten. Ansprechpartner waren dabei jeweils die Ländervertreter als Mitglieder der Projektgruppe 'Innovationen im Bildungswesen' in der Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung.

Auf Grund der Tatsache, dass nicht alle angeforderten Modellversuchsdokumentationen geeignet waren (s. u.), können nur sechs Modellversuchsergebnisse veröffentlicht werden. Diese werden nachfolgend mit Hilfe der Aspekte 'Allgemeine und strukturelle Angaben', 'Modellversuchsaufbau und –aktivitäten' sowie 'Modellversuchsbegleitende Transferaktivitäten/Fortbildungen/Tagungen' dargestellt und auf der Metaebene evaluiert.

Der Rücklauf der angeforderten Unterlagen stellt sich wie folgt dar:

Ein Bundesland gab keine Rückmeldung. Ein Modellversuch war in einem Bundesland unbekannt, wird aber von der BLK in der Förderliste geführt. Von vier Modellversuchen existieren keinerlei Berichte mehr und von weiteren vier sind nur Abschlussberichte bekannt, Sach- oder Zwischenberichte sind nicht existent oder un-

36) Die vollständige Auflistung der dem Förderbereich 'Berufliche Bildung' durch die BLK zugeordneten Modellversuche ist bei der BLK-Geschäftsstelle erhältlich.

37) Die in den Anhang (Tab. 4 - Anhang Teil B : Ausgewählte Modellversuche der Gruppe G4) eingestellte Auflistung von Modellversuchen entspricht der Auswahl derjenigen Modellversuche aus dem Förderbereich 'Berufliche Bildung', die auf Grundlage ihrer Betitelung der Gruppe G4 – Innovationen in bestehenden Fächern oder Schulformen zugeordnet sind und demnach Gegenstand der metaevaluativen Untersuchung sein sollen.

Die Übersicht gibt einerseits den Titel des Modellversuchs sowie dessen Laufzeit wieder, andererseits ist auch die Förderkennziffer angegeben, die durch die BLK und das BMBF vergeben werden. Diese Angaben sind für die Nachvollziehbarkeit des Vorgehens erforderlich, da nur mit so die Modellversuchsdokumentationen angefordert werden können. Nach Aussage des BMBF und der BLK liegen die Modellversuchsanträge, die weitergehende Einblicke geben könnten, nicht mehr vor.

auffindbar. Darüber hinaus heben die Berichte eines Modellversuches hervor, dass keine wissenschaftliche Begleitung vorhanden war und bei zwei anderen Modellversuchen hat die wissenschaftliche Begleitung nach eigenen Aussagen keine Evaluation vorgenommen. Die verbleibenden und für metaevaluative Zwecke geeigneten sechs Erprobungsvorhaben wurden in der Zeit zwischen 1991 und 2000 durchgeführt.

Die folgende Übersicht fasst den Rücklauf zusammen (n=19):

▪ Keine Rückmeldung	ein Bundesland
▪ Modellversuch 'unbekannt'	ein Modellversuch
▪ Keine Berichte vorhanden	vier Modellversuche
▪ Nur Abschlussberichte vorhanden	vier Modellversuche
▪ Keine wissenschaftliche Begleitung im Modellversuch	ein Modellversuch
▪ Wissenschaftliche Begleitung führt keine Evaluation durch	zwei Modellversuche
▪ Berichte sind für Meta-Evaluation geeignet	sechs Modellversuche

6.2 Darstellung und Evaluation des Modellversuchs KUNO

6.2.1 Beschreibung des Modellversuchs

6.2.1.1 Allgemeine und strukturelle Angaben

Der Modellversuch 'Kundenorientierung in der kaufmännischen Berufsausbildung' ist im Bundesland Bremen vom 01.11.1997 bis 30.04.2000 durchgeführt worden. Verantwortlich für Modellversuche im Land Bremen ist die Abteilung Curriculumentwicklung und Innovationsförderung im Landesinstitut für Schule (LIS), das dem Senatsressort für Bildung und Wissenschaft des Landes Bremen unterstellt ist (ZB 1999: 3, AB 2001: 1).

Beteiligt an dem Modellversuch waren insgesamt vier Schulstandorte im Land Bremen mit insgesamt 18 Projektgruppen, die in Vollzeit- und Teilzeitklassen arbeiten:

Zwei Schulzentren des Sekundarbereichs II bilden im dualen Bereich die Berufe Bank-, Versicherungs-, Industrie-, Automobil- und Bürokaufmann sowie Kaufmann für Bürokommunikation, Kaufmann im Einzelhandel, Handelsassistent, Dienstleistungsfachkraft im Postwesen, Sozialversicherungsfachangestellte und Fachangestellte für Arbeitsförderung aus. Außerdem werden in vollzeitschulischer Form eine berufsqualifizierende Maßnahme und die Ausbildung zum Wirtschaftsassistenten - Datenverarbeitung – und die einjährige sowie zweijährige Handelsschule angeboten. Darüber hinaus war ein Berufskolleg der Stadt Köln mit einem eigenständigen Modellversuchsantrag beteiligt.

Aus dieser Vielzahl an beteiligten Modellversuchsgruppen ergab sich eine hohe Bandbreite an Bildungsgängen und Lernfeldern, so dass nicht alle betroffenen Schüler mit allen Aspekten des Themas Kundenorientierung in Berührung kommen

sollten. Vielmehr lernten die unterschiedlichen Klassen wenige, z. T. sehr spezifische Facetten des Themenbereichs kennen (vgl. ZB 1999: 45).

Die wissenschaftliche Begleitung ist vom Institut Technik und Bildung (ITB), genauer der Abteilung Lernen, Lehren und Organisation, der Universität Bremen wahrgenommen worden. An der WBL wirkten ein Universitätsprofessor sowie ein wissenschaftlicher Angestellter mit (vgl. ZB 1999, AB 2001: Allgemeine Angaben). Der Modellversuchsverlauf wurde von der Projektleitung gesteuert. Mitglieder hierin waren eine Mitarbeiterin des LIS (s. o.) sowie ein schulfachlicher Mitarbeiter, die in enger Abstimmung mit der wissenschaftlichen Begleitung für die Konzeption der Modellversuchsinhalte und deren Umsetzung verantwortlich waren (ZB 1999: 7).

Dem Arbeits- und Zeitplan ist zu entnehmen, dass der Modellversuchszeitraum in drei Phasen unterteilt war: In der 'Konzeptions- und Gestaltungsphase' (Zeitdauer - vier Monate) konstituierte sich die Modellversuchsarbeitsgruppe, nahm Kontakte zu Institutionen und Ausbildungsbetrieben auf, vergab Arbeitsthemen usw. Die anschließende 'Realisierungs- und Spezifikationsphase' (Zeitdauer – neunzehn Monate), diente der Entwicklung, Erprobung und Revision von Lernsequenzen mit aktiven Lernmethoden, der Erstellung, Erprobung und Evaluation von Lehrmaterialien u. ä. Den Abschluss bildete die 'Auswertungs,- Implementations- und Transferphase' (Zeitdauer – sieben Monate), in die die 'Prüfung der Übertragbarkeit auf nachfolgende Klassen sowie andere Bildungsgänge und Formulierung entsprechender Empfehlungen', sowie die Auswertung des Modellversuchs, die auf Grund der Ausrichtung der WBL als kontinuierlicher Prozess angelegt und in der Hauptphase schon begonnen wurde, eingeordnet ist (vgl. ZB 1999: 8f).

6.2.1.2 Modellversuchsziele und –inhalte/Tätigkeiten der WBL auf Modellversuchsebene

Ziel des Modellversuchs war es, zentrale Aspekte der Kundenorientierung in den wirtschaftsberuflichen Fächern bereits in der kaufmännischen Erstausbildung bzw. in den berufsvorbereitenden Fachschulen zu erproben:

"Das Globalziel kann im Kontext der kaufmännischen Berufsbildung als Entwicklung einer umfassenden kundenorientierten Handlungskompetenz verstanden werden" (ZB 1999: 23).

Gegenstand des Modellversuchs war die Entwicklung von Unterrichtseinheiten zum Themenbereich Kundenorientierung und die Erfassung ihrer Wirkungen bei den Schülerinnen und Schülern sowie die Qualifizierung der beteiligten Lehrkräfte zur Ausbildung der notwendigen Kompetenzbereiche s. u. (vgl. ZB 1999: 46). Wegen der sinkenden Halbwertszeit des Wissens und der Notwendigkeit der Ausbildung von Fach-, Methoden- und Sozialkompetenz sollen Methoden des selbstorganisierten Lernens zum Einsatz kommen, um die angestrebten Kompetenzbereiche auszubilden (vgl. ZB 1999: 31):

Fachliche Sicherheit, Kommunikation und Servicetugenden als Kompetenzbereiche kundenorientierten Handelns sollten dabei durch die Vermittlung der nachfolgenden Lernbereiche entwickelt werden - Kundenorientierung im Marketing, Ge-

schäftsprozessorganisation/Informationssysteme, Kommunikationsformen Kunde-Unternehmen, Kundenorientierung als Thema im Deutschunterricht, Kaufvertragsgestaltung und Beschwerdemanagement und Kundenorientierung im Finanzdienstleistungsbereich.

Ausgehend von der Präzisierung und Differenzierung des Begriffs ‚Kundenorientierung‘ in Form der Entwicklung von Leitbildern und Leitlinien erfolgt deren Operationalisierung im Hinblick auf die didaktisch-methodische Umsetzung, d. h. es wurden Handlungskompetenzen für die unterschiedlichen kaufmännischen Bildungsgänge und Lernfelder herausgearbeitet und formuliert; ganzheitlich organisierte Bildungsprozesse sollten die Entwicklung der Handlungskompetenzen unterstützen (vgl. ZB 1999: 22ff).

Darüber hinaus war es ein Ziel, im Rahmen des Modellversuchs Möglichkeiten zu eruieren, Ausbildungsinhalte zur Kundenorientierung als Ausgangspunkt für Lernortkooperationen aufzugreifen.

Die wissenschaftliche Begleitung sah ihren Schwerpunkt im Rahmen der Beratungsfunktion in der "[...] Beratung der Projektmitarbeiterinnen bei der Entwicklung der Unterrichtseinheiten unter besonderer Berücksichtigung des Konzepts des selbstorganisierten Lernens" (ZB 1999: 31). In 'Workshops' wurden die Projektmitarbeiter daher mit dem Ansatz und den Methoden des selbstorganisierten Lernens vertraut gemacht.

Die Evaluationsfunktion der WBL sollte eine formative und summative Ausprägung erhalten, wobei die formative Evaluation mit ihren Steuerungs- und Optimierungsfunktionen, bezogen auf die Modellversuchsdurchführung, im Vordergrund gestanden hat. "Die angestrebten praxisbezogenen und handlungsrelevanten Ergebnisse sollen zur Optimierung der Bildungsmaßnahme beitragen. Formative Evaluation ist daher [...] mit Verfahren der Qualitätssicherung zu vergleichen" (ZB 1999: 30). Im Gegensatz zur summativen Evaluation impliziere sie die Beratung, da die Erkenntnisse aus der formativen Evaluation in den laufenden Modellversuchsprozess den Schulpraktikern zurück gespiegelt werden können.

Als Evaluationsmethoden auf MV-Ebene sind teil-standardisierte Beobachtungsbögen im Rahmen eines Telefontrainings sowie Fragebögen für externe Übungsfirmen, die mit einer zu evaluierenden Übungsfirma in Kontakt stehen, zum Einsatz gekommen. Darüber hinaus wurden mit einigen Klassen Einzel- oder Gruppeninterviews geführt. Als methodische Herangehensweise dienten hier narrative Interviews. Um das bei den Schülern vorherrschende Verständnis von Kundenorientierung zu erfassen, wurden von der WBL entwickelte Fragebögen eingesetzt (vgl. ZB 1999: 46).

Da neben der Entwicklung der Unterrichtskonzepte auch die Qualifizierung der beteiligten Lehrkräfte zu den Zielen des Modellversuchs zählte, sollten die kognitiven, motivationalen und emotionalen Wirkungen der Fortbildungsveranstaltungen sowie die Erfahrungen bei der Entwicklung und Umsetzung der Unterrichtseinheiten

ten und den 'erlebten Modellversuchsarbeiten' mit Hilfe leitfadengestützter Einzelgespräche evaluiert werden (vgl. ZB 1999: 46; AB 2001: 390).

6.2.1.3 Fortbildungen/Tagungen/modellversuchsbegleitende Transferaktivitäten

Im Rahmen des Modellversuches diente den Mitgliedern der Projektgruppe, der wissenschaftlichen Begleitung sowie der Projektleitung eine Auftaktveranstaltung als 'Orientierungsphase'. Darüber hinaus fanden Veranstaltungen zum Telefontraining und zum Verkaufsgespräch (drei Tage), Selbstorganisierten Lernen (eins Komma fünf Tage) und zu Präsentationstechniken (zwei Tage) statt (vgl. ZB 1999: 99ff).

Die inhaltlichen und methodischen Erkenntnisse flossen außerdem in eine Lehrveranstaltung für den Studiengang Lehramt Sekundarstufe II mit beruflicher Fachrichtung Wirtschaft, in drei schulinterne Lehrerfortbildungen, sowie in eine erste und eine zweite Staatsexamensarbeit ein (vgl. AB 2001: 328f).

6.2.1.4 Ergebnisse des Modellversuchs aus Sicht der WBL

Im Abschlussbericht wird zum Gliederungspunkt 'Ergebnissen, Perspektiven und Empfehlungen' einleitend hervorgehoben, dass die Ergebnisse des Modellversuchs (Objektebene) mit den formulierten Projektzielen kritisch konfrontiert werden sollen (vgl. AB 2001: 324). Diese Darstellung unterbleibt jedoch.

Im Hinblick auf die Förderung der Qualifikation 'Kundenorientierung' als Ergebnis ihrer Evaluation (Evaluationsebene) kommt die WBL zu folgenden Ergebnissen:

- Die Schüler der Kontrollgruppe, die in die Evaluation einbezogen waren, hätten ein Verständnis von Kundenorientierung, das auf ihren Erfahrungen als Kunde beruhe und durch fehlende 'Ganzheitlichkeit' gekennzeichnet sei.
- Die in den Modellversuch involvierten Schülern 'scheinen' nach Angaben der WBL für das Thema 'Kundenorientierung' sensibler zu sein.
- Der 'Grad der Sensibilisierung' korreliere mit dem Stundenumfang der KUNO-Unterrichtseinheiten und den KUNO-Lehrern pro Klasse.
- Soweit die involvierten Schüler dies beurteilen könnten, scheine das Thema Kundenorientierung in den Ausbildungsunternehmen 'kein Thema' zu sein.
- Die entwickelten Unterrichtseinheiten würden überwiegend positiv eingeschätzt.
- Der Institution Schule werde überwiegend eine geringe Kundenorientierung zugeschrieben (vgl. AB 2001: 402f).

6.2.2 Darstellung des summativen Evaluationsansatzes der WBL im Modellversuch KUNO

Im Hinblick auf die Ergebnisse der summativen Evaluation hat die WBL herausgestellt, dass die Heterogenität der entwickelten Unterrichtseinheiten und deren Einsatz in unterschiedlichen Schulen und Klassen die homogene Beurteilung eines für alle beteiligten Schüler verbindlichen und vergleichbaren Modellversuchsergebnisses insgesamt verhindere (vgl. ZB 1999: 45).

Gleichwohl wies sie an anderer Stelle darauf hin, dass die summative Evaluation eine abschließende und verallgemeinerbare Beurteilung von Maßnahmen anstrebe,

mit der der Erfolg oder der Nutzen einer Maßnahme bewertet werden solle (vgl. ZB 1999: 29). Ob es sich bei der letztgenannten Aussage um eine allgemeine Feststellung bezüglich der Funktion summativer Evaluationen handelt oder eine Absichtserklärung darstellt oder als Widerspruch zur o. g. Aussage angesehen werden muss, kann den Dokumentationen nicht entnommen werden.

Als Methode der summativen Evaluation kamen leitfadengestützte Einzelgespräche zum Einsatz, an denen insgesamt 18 Probanden teilnahmen. Sechs von ihnen waren nicht Schüler von Klassen, die entsprechend der Modellversuchsintention unterrichtet wurden, sondern dienten als Mitglieder einer Kontrollgruppe.

Bei den Befragten handelte es sich um Jugendliche bzw. junge Erwachsene im Alter von 17 bis 25 Jahren (sieben weiblich, elf männlich), von denen neun Schüler die Höhere Handelsschule, sieben eine Berufsausbildung im dualen System (keine genauere Spezifizierung) und zwei Schüler eine Ausbildung zum Wirtschaftsassistenten durchlaufen haben. Die Auswahl der Schüler erfolgte zum Teil auf freiwilliger Basis, zum Teil durch Bestimmung durch den jeweiligen Lehrer.

Die Interviews dauerten zwischen 30 und 90 Minuten (Durchschnitt 55 Minuten), wurden aufgezeichnet und anschließend transkribiert (vgl. AB 2001: 390f).

Der Interviewleitfaden umfasste drei Themenbereiche: 'Kundenorientierung allgemein', 'Kundenorientierung im (Auslands-) Unternehmen', 'Kundenorientierung in der Schule'. Während im ersten Fragenbereich das allgemeine Verständnis von Kundenorientierung erfasst werden sollte, wurde der Fokus im zweiten Bereich auf die Kundenorientierung in der 'Praxis' gerichtet. Der dritte Fragenblock bezog sich bei den Schülern der Untersuchungsgruppe auf die konkreten Unterrichtsvorhaben im Rahmen des Modellversuchs (Inhalte, Vermittlung, Frage, ob und in welcher Form Kundenorientierung im Unterricht thematisiert wurde) und kann daher als summativer Evaluationsgegenstand im engeren Sinne betrachtet werden. Die Kontrollgruppe wurde in diesem Zusammenhang dahingehend befragt, ob und in welchem Umfang Kundenorientierung im Unterricht thematisiert wurde.

Die Auswertung der Befragung stellt die teilweise Wiedergabe der Interviews (zwischen sechs und neun Antwortauszüge), ergänzt durch kurze Kommentare, dar (vgl. AB 2001: 393).

6.2.3 Ergebnis der Untersuchung - Ausprägungsgrade der Evaluationsstandards

S	N1	N3	N4	N5	N	D1	K1	K5
A	1	0	0	0	0,25	0	0	0

S	K6	K7	K	G1	G2	G3	G4	G5
A	3	0	0,75	1	1	0	1	1

S	G6	G8	G9	G10	G12	G
A	0	-	0	0	0	0,44

S=Standard; A=Ausprägungsgrad

Tab. 5: Ergebnis der metaevaluativen Untersuchung MV KUNO/Bremen

N1 – Ermittlung der Beteiligten und Betroffenen: Im Rahmen der allgemeinen Angaben werden die Projektleitung sowie die Mitglieder der wissenschaftlichen Begleitung genannt. Auch die Mitglieder der Projektgruppe (Lehrer) werden namentlich und entsprechend ihrer Zugehörigkeit zu unterschiedlichen Schulen aufgelistet (vgl. ZB 1999; AB 2001: Allgemeine Angaben) und die betroffenen Klassen werden ausgewählt (vgl. ZB 1999: S. 48ff). Aus den beteiligten Schulen werden 18 Schüler als Mitglieder der Untersuchungsgruppe der summativen Evaluation (12 Schüler besuchen Modellversuchsklassen, 6 Schüler dienen als Kontrollgruppe, stammen daher aus modellversuchsfremden Klassen) bestimmt, die als Probanden für die narrativen Interviews dienen. Diese werden hinsichtlich ihres Geschlechts, ihres Alters und der von ihnen besuchten Schulformen benannt. Potenzielle Beteiligte und Betroffene, wie z. B. Einzelhandelsverbände oder andere Zusammenschlüsse von Interessengruppen, Mitglieder möglicher Lernortkooperationen und die entsprechenden Informationsbedürfnisse werden nicht identifiziert, so dass eine Gruppierung der B&B nach dem Kriterium der Einflussmöglichkeit unterbleiben muss (AB 2001: 390f).

N3 – Umfang und Auswahl der Informationen: Die WBL kommt auf Grund ihrer Befragung zu dem Schluss, dass die involvierten Schüler nur 'ausschnittthafte Einblicke' in den Themenbereich Kundenorientierung erhalten haben. Aussagen über das Ausmaß der Zielerreichung sind den Dokumentationen nicht zu entnehmen, lediglich generelle Einschätzungen sind vorhanden: "Erwartungsgemäß war das Verständnis von Kundenorientierung bei der Untersuchungsgruppe, also den Schülern/innen, die in den Modellversuch involviert waren, in den meisten Fällen umfassender bzw. differenzierter als bei der Kontrollgruppe" (AB 2001: 396). Das Fazit aus den Befragungen führt daher auch nur zu Aussagen, deren Wortwahl den Erkenntnissen den Charakter von Vermutungen verleiht. Hier werden vor allem folgende Adverben und Verben verwandt: 'Das Verständnis von Kundenorientierung beruht offenbar "[...], die in den Modellversuch involvierten Schüler/innen scheinen [...] sensibilisierter zu sein" (AB 2001: 402f). Diese wenig konkreten Aussagen können nicht als Herausheben 'bedeutender Informationen aus dem MV' angesehen werden.

N4 – Feststellung von Werten: Die Prinzipien der WBL im Hinblick auf die Evaluation des Vorgehens auf der Objektebene (formative Evaluation) werden dargestellt (vgl. ZB 1999: 44f; AB 2001: 388). Darüber hinaus erläutert die WBL den von ihnen in den Modellversuch eingebrachten Ansatz des 'selbstorganisierten Lernens nach Straka', so dass die für die Evaluation geltenden Wertmaßstäbe hieraus zumindest im Ansatz entnommen werden können. Zielformulierungen oder explizite Absprachen mit Unternehmensvertretern oder anderen Beteiligten und Betroffenen hinsichtlich einer Wertzuordnung für die Ergebnisse der summativen Evaluation, die eine abschließende Einschätzung des Vorgehens und der Interpretationen, auch durch potenzielle Transferinteressenten, ermöglichen würden fehlen jedoch.

N5 – Klarheit des Berichts: Es wird unterlassen, die Ziele der Evaluation darzustellen (siehe G3). Die eingesetzte Form der Befragung wird beschrieben und die

Antworten werden teilweise aufgelistet. Eine strukturierte Auswertung, die sich zu den Evaluationszielen in Verbindung setzen könnte muss so unterbleiben (s. o.). Die Schlussfolgerungen können sich demnach nicht auf eine abgesicherte Grundlage beziehen. Die WBL kommt zu der Erkenntnis: "Die Struktur des Modellversuchs erschwert eine abschließende Evaluation im Sinne abschließender und verallgemeinerbarer Ergebnisse" (AB 2001: 394). Diese Aussage ist nicht geeignet, den Modellversuchsaufbau in einen Planungsmodellversuch einzubeziehen, wie es dieser Standard generell impliziert.

Resümee Standardbereich N:

Der Standardbereich N (Nützlichkeitsstandards) liegt hinsichtlich seiner durchschnittlichen Ausprägung im Bereich des schwächsten Ausprägungsgrades der verwendeten Skala. D. h. die Modellversuchsevaluation richtet sich nur in einem Maße nach den Informationsbedürfnissen der B&B, das zwischen der Nichtbeachtung des Standards und der Berücksichtigung zu einem geringen Teil liegt. Zurückzuführen ist dieses Ergebnis auf die Nichterfüllung der Standards N3, N4 und N5, da u. a. Angaben über den Verlauf und die zu Grunde gelegten Ziele der Evaluation bzw. der Beratertätigkeit während des MV-Verlaufes nicht erfolgen.

D1 – Praktische Verfahren: Die zum Einsatz kommende Methode der Befragung (leitfadengestützte Einzelgespräche) ist im Hinblick auf die Anzahl der Befragten und der Auswertung zeitlich und hinsichtlich des Aufwandes nicht derart umfangreich, als dass diese nicht vertretbar erscheint (12*55 Minuten=siehe 6.2.2). Allerdings ist es nicht angebracht, einen Kompetenzzuwachs, der durch Methoden des selbstorganisierten Lernens herbeigeführt werden soll (s. o.), durch eine Befragung zu ermitteln, die hauptsächlich auf das Feststellen von Empathie im Hinblick auf den Lerngegenstand 'Kundenorientierung' ausgerichtet ist (Korrelation zu G5; vgl. AB 2001: 390ff).

K1 – Unterstützung der Dienstleistungsorientierung: Es ist Aufgabe einer Evaluation, beabsichtigte und unbeabsichtigte Wirkungen des Programms auf die Lernenden zu erkennen (vgl. Sanders 1999: 109) und Transparenz herzustellen. Diese Art der 'Stärken- und Schwächenanalyse' ist den Dokumentationen nicht zu entnehmen. Funktion der Evaluatoren im Sinne dieses Qualitätsstandards ist es jedoch, über die Interessen von Pädagogen und Organisationen hinaus zu blicken, die Entwicklung der Lernenden zu verbessern und als Ergebnis der Evaluation Informationen zu generieren, die für Planungsmodellversuche bzw. bildungspolitische Entscheidungen genutzt werden können. Ziel ist es, die Informationsbedürfnisse der Beteiligten und Betroffenen zu befriedigen (Dienstleistungsfunktion). Nach Lage der Berichte sind die Erkenntnisse demnach nur sehr bedingt geeignet, die Anforderungen des Standards zu erfüllen.

K5 – Vollständige und faire Einschätzung: Eine bewertete Auflistung von Stärken und Schwächen des Modellversuchs unterbleibt, was auf Grund der Nichtbeachtung der Standards N5 (Klarheit des Berichts) und K1 (Unterstützung der Dienstleistungsorientierung) eine zwingende Folge darstellt. Selbst beabsichtigte

Erkenntnisse im Sinne von Ergebnissen des Modellversuchs werden nicht dargestellt (siehe auch Kap. 6.2.1.4).

K6 – Offenlegung der Ergebnisse: Der Zwischenbericht und Abschlussbericht sind über den Senator für Bildung und Wissenschaft in Bremen zu beziehen. Auch werden der Projektleiter und die beteiligten Lehrer genannt. Darüber hinaus enthält der Abschlussbericht die Hinweise auf zwei Staatsexamensarbeiten sowie diverse zum Modellversuch erschienene Zeitungsartikel (vgl. AB 2001: 6).

K7 – Deklaration von Interessenkonflikten: Auf mögliche Interessenkonflikte wird nicht hingewiesen, auch fehlen ex-ante formulierte Schutzmaßnahmen oder die Angabe über Vorbesprechungen mit den anderen Modellversuchsbeteiligten hierzu. Wegen der Vielzahl von beteiligten Schulen und Bildungsgängen sowie der Bandbreite der Interessenten an den (je speziellen) Modellversuchsergebnissen ist jedoch das Entstehen von Konflikten persönlicher oder finanzieller Art wahrscheinlich.

Resümee Standardbereich K:

Der durchschnittlich erreichte Ausprägungsgrad der Korrektheitsstandards bewegt sich im unteren Viertel der Ausprägungsgrade. Dass dieser Standardbereich nicht insgesamt als 'nicht beachtet' eingeschätzt werden muss, ist auf den Standard K6 zurückzuführen, der die Offenlegung sämtlicher Dokumentationen, Erklärungen u. ä. fordert und im vorliegenden Fall den Ausprägungsgrad ,3' erreicht, da die notwendigen Hinweise auf Bezugsquellen für die Dokumentationen und weitere Veröffentlichungen enthalten sind (vgl. AB 2000: 6). Die WBL ist demnach lediglich ihren dokumentarischen Pflichten nachgekommen; dem Wohlergehen der Beteiligten und Betroffenen sowie deren Interessen (Korrelation zu N1) wird keine Aufmerksamkeit geschenkt.

G1 – Modellversuchsdokumentationen: Dem Zwischenbericht ist ein Arbeitsplan zu entnehmen (ZB 1999: 8f), der einen groben Überblick über den Modellversuchsverlauf erlaubt. Weitere Angaben, die über die ursprüngliche Modellversuchsplanung und die notwendigen Änderungen informieren, sind in den Dokumentationen nicht enthalten. Der Ansatz der formativen Evaluation, wie er von der WBL auch als Beratungsfunktion angesehen wird, hätte Änderungen hervorbringen und zur Modifizierung beitragen sollen (vgl. ZB 1999: 45).

G2 – Kontextanalyse: Im Zwischenbericht wird die derzeitige marktwirtschaftliche Situation unter dem Stichwort ,Servicewüste' Deutschland dokumentiert (vgl. ZB 1999: 11ff). Das weitere politisch und soziale Umfeld wird jedoch ebenso wenig aufgezeigt, wie mögliche parallel laufende konkurrierende Aktivitäten. Dem Arbeitsplan ist jedoch zumindest eine zeitliche Dimensionierung zu entnehmen. Auch der Hinweis auf die Aktivitäten am zweiten MV-Standort werden erwähnt und die beteiligten Lehrer werden entsprechend ihrer Zugehörigkeit zu den Schulstandorten genannt (siehe ZB/AB: allgemeine Angaben).

G3 – Beschreibung von Zielen und Vorgehen: Es erfolgt keine Nennung der vorgesehenen Evaluationsziele, so dass Erklärungen über beabsichtigten Nutzungen vor, während und nach Abschluss des Modellversuchs unterbleiben müssen. Eine begründete Beschreibung des Evaluationsvorgehens unterbleibt außerdem. Lediglich im Abschlussbericht erfolgt eine kurze Erläuterung hinsichtlich der Durchführung der leitfadengestützten Interviews (vgl. AB 2001: 390).

G4 – Verlässliche Informationsquellen: Die für die Befragung ausgewählten Schüler werden hinsichtlich ihres Alters und Geschlechts beschrieben; über die Auswahl der Probanden erfolgt nur der Hinweis, dass diese 'zum Teil' auf freiwilliger Basis ausgewählt wurden. Das Problem der Beeinflussung der Befragungsergebnisse durch die Probanden wird in den Berichten nicht angesprochen. Ergänzende Instrumente der Informationsgewinnung, die den Nachteil der Methode der Befragung ausgleichen, kommen nicht zum Einsatz. Angaben über die Grundgesamtheit und mögliche Fluktuationen fehlen außerdem. Die der Befragung zu Grunde liegenden Interviewimpulse sind dokumentiert, jedoch nicht begründet worden.

G5 – Valide Informationen: Eine im Hinblick auf die Validität der eingesetzten leitfadengestützten Einzelgespräche ausgerichtete Begründung unterbleibt ebenso, wie eine bewertende Beschreibung der gesammelten Informationen. Da die Ziele der Evaluation nicht dargestellt werden, können die gesammelten Ergebnisse nicht auf diese bezogen werden.

Darüber hinaus scheinen Probandenbefragungen im Hinblick auf das Globalziel auf der Objektebene 'Ausbildung von Handlungskompetenz' nur bedingt geeignet (siehe Standard D1), eine Einschätzung dieser Zielerreichung vorzunehmen, da keine vorher-nachher-Untersuchung mit Ermittlung der Qualifikationsentwicklung durchgeführt wurde. Schlussfolgerungen, die sich aus der Befragung auf die eingesetzten methodischen Mittel (SOL-Konzept) beziehen und deren Eignung in Bezug auf die Entwicklung der Qualifikation 'Kundenorientierung' einschätzen lassen fehlen außerdem.

G6 – Reliable Informationen: Leitfadengestützte Interviews sind von ihrem Charakter (Offenheit des Interviews) her in besonderem Maße personen- und situationsabhängig, so dass die Schlussfolgerungen aus diesem Untersuchungsschritt auf Grund dieser Störeinflüsse und durch die Deutungen der Informationen durch den Interviewer durch andere Untersuchungsmethoden minimiert werden müssten. Hinweise auf das Erkennen dieses Problem durch die WBL sind den Berichten nicht zu entnehmen. Zwar handelt es sich bei den hier vorliegenden Störeinflüssen um 'systematische Stör-Faktoren', da diese jedoch nicht ausgeglichen werden (keine anderen Evaluationsmethoden), ist die Reliabilität der Informationen, die diese Untersuchung hervorbringt, in Frage zu stellen (vgl. auch Sanders 1999: 183).

G8 – Analyse quantitativer Informationen: entfällt, da keine quantitative Daten erhoben werden.

G9 - Analyse qualitativer Informationen: Die WBL greift in ihrer Evaluation nur auf eine Quelle (Schüleraussagen) zurückgreift und lässt die Ergebnisse nicht einer unabhängigen Überprüfung unterziehen. Aussagen aus den Interviews werden Frage-Kategorien (Kundenorientierung allgemein, im Unternehmen und in der Schule) zugeordnet (vgl. AB 2001: 391ff), indem Auszüge aus den Interview-Bereichen abgedruckt werden. Nach welchen Kriterien die Auszüge ausgewählt wurden, ist den Berichten nicht zu entnehmen. Schlussfolgerungen werden keine gezogen; es sind lediglich Empfehlungen vorhanden, die nicht auf das eingesetzte SOL-Konzept zurückgeführt werden. Die gesammelten Informationen können nicht auf gesetzte Ziele bezogen werden und sind daher nicht geeignet, Ausgangsfragen zu beantworten.

G10 – Begründete Schlussfolgerungen: Es werden Empfehlungen abgedruckt, die nicht als Schlussfolgerungen im Sinne von Aussagen, die durch Rückführung auf Ursachen nachvollziehbar sind bezeichnet werden können (vgl. AB 2001: 402). Da die Beteiligten und Betroffenen nur unzureichend ermittelt wurden und somit entsprechende Aussagen über deren Informationsbedürfnisse ebenso fehlen, wie Untersuchungsmethoden, die die Eignung der Methoden zur Zielerreichung (SOL-Konzept) auf der Objektebene einschätzen lassen, sind die Aussagen der WBL im Hinblick auf die Ergebnisse ihrer Evaluation insgesamt wenig substantiell.

G12- Metaevaluation: Angaben zu einer möglichen Metaevaluation sind in den Berichten nicht enthalten.

Resümee Standardbereich G:

Die Genauigkeitsstandards erreichen die geringste durchschnittliche Ausprägung der verwendeten Standardbereiche mit mehreren eingesetzten Evaluationsstandards. Lediglich die Standards G1 (MV-Dokumentationen) und G2 (Kontextanalyse) weichen vom Ausprägungsgrad '0' ab, da die Dokumentationen zum einen einen Arbeitsplan enthalten, dem auch eine zeitliche Dimensionierung zu entnehmen ist, zum anderen ist eine recht ausführliche Schilderung der marktwirtschaftlichen Situation vorhanden, die zur Initiierung des Modellversuchs Anlass gab (ZB 1999: 8f) Darüber hinaus können oberflächliche Informationen zur befragten Personengruppe den Dokumentationen entnommen werden. Alle anderen Evaluationsstandards werden nicht erfüllt, so dass die 'Bewährung der Modellversuchsmaßnahmen' nicht nachgewiesen werden kann und keine Informationen generiert werden, die als Schlussfolgerungen im Sinne von 'Ergebnis-Ursachen-Rückführung' bezeichnet werden können.

Die Anwendung des Kongruenzschemas zur Feststellung der gegenseitigen Abhängigkeit von Modellversuchsaufbau inkl. Evaluationsverfahren und Verfahren der Zielermittlung konnte nicht durchgeführt werden, da die notwendigen Ziele der Evaluation (Prüfaspekt 1) nicht vorhanden sind.

Zwischenergebnis – Analyse und Interpretation der Ergebnisse:

Die schwache Ausprägung der Nützlichkeitsstandards macht deutlich, dass sich die Evaluatoren nicht ausführlich mit ihren Adressaten vertraut gemacht, ihre Informationsbedürfnisse ermittelt und eine auf diese Bedürfnisse zugeschnittene Evaluation geplant haben, so dass die Analyse und Bewertung der Ergebnisse der Untersuchung Erkenntnisse hervorbringt, die nicht zielgruppenadäquat sein können. Wenn sie für den Auftraggeber und die anderen Betroffenen und Beteiligten von Interesse sind, dann handelt es dabei um eine zufällige Übereinstimmung von Informationsbedürfnissen und Ergebnissen der Evaluation.

Die Mängel bei der Berücksichtigung der Nützlichkeitsstandards zeigen sich auch bei der Ausprägung der Korrektheitsstandards. K1 (Unterstützung der Dienstleistungsorientierung) kann nicht erfüllt werden, da die Standards N1 (Ermittlung der Beteiligten und Betroffenen) sowie N3 (Umfang und Auswahl der Informationen) nicht oder nur zu einem sehr begrenzten Maße erfüllt werden. Mit Ausnahme des Standards K6 (Offenlegung der Ergebnisse), der sich im Wesentlichen auf die Dokumentationspflicht, die auch durch die BLK gegeben ist, bezieht, sind alle anderen Standards nicht erfüllt worden. Obwohl die WBL die Bedeutung der formativen Evaluation hervorhebt (ZB 1999: 45), sind keinerlei Aussagen über den geplanten Verlauf des Modellversuchs und mögliche Änderungen, die sich auf Grund bestimmter Schwächen ergeben haben in den Dokumentationen enthalten. Das Wohlergehen der Beteiligten und Betroffenen wird durch die Nichtbeachtung dieses Standardbereiches missachtet, die Rechte anderer in Bezug auf die Informationen bleiben unbeachtet.

Durch die Korrelation der Wirkungen des Standards D1 (Praktische Verfahren) mit den Genauigkeitsstandards zeigen sich die Mängel bei der Wahl der Untersuchungsmethode auch hier: Mit Ausnahme von G1 (MV-Dokumentationen) und G2 (Kontextanalyse), die den Ausprägungsgrad 1 erreichen, sind alle anderen Standards nicht erfüllt. Aufgabe dieses Standardbereiches ist es sicherzustellen, "[...] dass eine Evaluation präzise Informationen über die Verwendbarkeit und/oder Güte eines Programms offenlegt und vermittelt" (Sanders 1999: 29).

Gerade die schwache Ausprägung des Standardbereiches der Genauigkeitsstandards macht deutlich, dass die Evaluation keine präzisen Informationen über die Güte des Modellversuchs hervorbringt.

6.3 Darstellung und Evaluation des Modellversuchs UMWELT

6.3.1 Beschreibung des Modellversuchs

6.3.1.1 Allgemeine und strukturelle Angaben

Der Modellversuch 'Umweltbildung in Berufen der öffentlichen Verwaltung – Entwicklung und Erprobung ökologischer Lernangebote in der Aus- und Weiterbildung' ist im Bundesland Bremen vom 01.01.1993 bis 31.12.1995 durchgeführt worden.

Als Modellversuchsschule fungierte die Verwaltungsschule der Freien Hansestadt Bremen. Bezogen auf den Charakter der Verwaltungsschule ist diese nicht öffentliche Schule im schulrechtlichen Sinne, sondern eine Schule der öffentlichen Verwaltung für die Ausbildung der 'Beamtenberufe': Mittlerer Steuerverwaltungsdienst, Mittlerer Justizdienst, Justizvollzugsdienst. Daneben zeichnet sie aber auch für verschiedene Berufe verantwortlich, die nach dem Berufsbildungsgesetz geordnet sind (Verwaltungsfachangestellte, Fachangestellte für Bürokommunikation, Assistenten in Bibliotheken), so dass eine Öffnung in Richtung 'öffentliche Berufsschule' stattgefunden hat. Dieser Charakterwechsel wird durch personalpolitische Entscheidungen unterstützt, wonach auf die Beamtenausbildung des mittleren allgemeinen Verwaltungsdienstes zugunsten der Verwaltungsfachangestelltenausbildung und anderen verwaltenden und bürotechnischen Ausbildungen verzichtet wird (vgl. ZB 1994: 7).

Daneben ist die Verwaltungsschule im Auftrag der Senatskommission für das Personalwesen für einen großen Teil der Fort- und Weiterbildung der Bediensteten der Freien und Hansestadt Bremen verantwortlich. Daher hat auch die Möglichkeit zu Beginn des Modellversuchs bestanden, durch Fortbildungen das Kollegium über die beabsichtigten Ziele zu informieren und umweltbezogene Inhalte sowie die Grundsätze handlungsorientierten Unterrichts zu vermitteln (vgl. AB 1996: 18).

Eine weitere Besonderheit der Schule ist die geringe Zahl der hauptamtlichen Lehrer (11), diese wurden jedoch vollständig in die Modellversuchsaktivitäten einbezogen; zum Teil traf dies auch auf die nebenamtlichen Lehrbeauftragten zu. Darüber hinaus wurde eine 'engere' Modellversuchsarbeitsgruppe (Projektleitung) initiiert, die aus drei schulischen Vertretern, zwei Mitarbeitern von Behörden der Hansestadt (behördenbezogener Projektleiter und Verantwortlicher für den praktischen Teil der Berufsausbildung) und zwei Mitgliedern der wissenschaftlichen Begleitung (Universität Bremen) bestanden hat.

Zur fachlichen Unterstützung wurde eine Kooperation der Modellversuchsarbeitsgruppe mit dem Institut für Umweltbildung im Beruf, Berlin, die im Rahmen von Werkverträgen tätig war, ins Leben berufen (vgl. ZB 1994: 12).

Die wissenschaftliche Begleitung erfolgte durch den Fachbereich 2 (Fachdidaktik Chemie) der Universität Bremen. An der WBL wirkten ein Universitätsprofessor und ein wissenschaftlicher Angestellter mit.

6.3.1.2 Modellversuchsziele und -inhalte/Tätigkeiten der WBL auf Modellversuchsebene

Ziel des Modellversuchs war die Entwicklung und Erprobung von Unterrichtseinheiten zum Themenbereich 'Umweltbildung', die dem didaktisch-methodischen Konzept der Handlungsorientierung verpflichtet sein sollten (vgl. AB 1996: 15), wobei auf eine integrierte, d. h. mit den übrigen fachlichen Inhalten verknüpfte fachbezogene Umweltbildung abgestellt wurde (vgl. AB 1996: 14).

Als Formalziel formulierten die Verantwortlichen, dass "[...] die zukünftigen Verwaltungsfachangestellten und Fachgestellten für Bürokommunikation als Personen zu befähigen (sind, Anm. Verf.), sich in ihrem Beruf umweltgerecht zu verhalten, d. h. sowohl bei der unmittelbaren Aufgabenerfüllung als auch im Umfeld der Aufgaben Umweltgesichtspunkte zu beachten" (AB 1996: 13f).

Ausschlaggebend hierfür war die Erkenntnis, dass die Verwaltung in doppelter Hinsicht auf die Umwelt einwirkt: einerseits ist sie als ‚größter Betrieb‘ der Hansestadt auch ein bedeutender ‚Ressourcenverbraucher und Umweltbelasteter‘, andererseits "[...] gestaltet die Verwaltung Umwelt – auf den unterschiedlichsten administrativen Feldern im Rahmen der rechtlichen Vorgaben und im Zusammenwirken mit der Politik" (AB 1996: 12).

Das Aufgabenfeld der WBL wurde von dieser mit neun Punkten charakterisiert. Die Aufgabenbeschreibungen reichen von der Mitarbeit im Hinblick auf die Bestimmung und Ausgestaltung der Unterrichtsgegenstände, der Entwicklung von Forschungsinstrumenten der wissenschaftlichen Begleitung für die Erprobungen bis hin zur Vorbereitung und Durchführung von Lehrerfortbildungen. Die Tätigkeiten wurden als 'konstruktiv-inhaltlich' gekennzeichnet, so dass der zum Einsatz kommende 'Begleitforschungsansatz' derart ausgerichtet sein musste, dass er den Beratungs- und Unterstützungscharakter der Tätigkeiten der WBL auf Modellversuchsebene fördern konnte; gemeint war die 'Aktionsforschung'.

Die gegenseitige Beeinflussung von Ausgestaltung, Durchführung und Analyse des Modellversuchs war hierbei ebenso beabsichtigt wie die Offenheit des Forschungsvorhabens, die sich in einem mehrstufigen Diskurs, der auch konträre Orientierungen zulässt, zeigte. Die Resultate der Auseinandersetzung flossen dann möglichst sofort in die Ausgestaltung des Modellversuchs ein, so dass der beabsichtigte Kreislauf mit wechselseitigen Einflussmöglichkeiten entstand (vgl. ZB 2 1995 Anhang: 1f). Dieses Vorgehen entspricht dem Grundgedanken der formativen Evaluation, da er als zyklisch angelegter Evaluationsprozess mit Rückkopplung beschrieben werden kann (vgl. Bortz 1995: 107).

Entsprechend hob die wissenschaftliche Begleitung die Vorläufigkeit des Verständnisses der zu untersuchenden Gegebenheiten im Modellversuch zu Beginn ihrer Tätigkeiten hervor, "[...] so dass der Untersuchungsgegenstand erst im Forschungsprozess ausformuliert werden kann. Die eigentlichen Untersuchungsschwerpunkte bilden sich somit erst im Verlauf des Forschungsprozesses [...]" (ZB 2 1995 Anhang: 2f).

Die eingesetzten Evaluationsmethoden sind dabei so ausgewählt bzw. gestaltet worden, dass eine 'unmittelbare Beziehung' zu den Forschungssubjekten hergestellt wurde: Befragungen, Interviews, Beobachtungen der Beteiligten und Betroffenen am Arbeitsplatz. "Der Forscher tritt selbst in das zu untersuchende Feld ein und bietet somit mannigfaltige Möglichkeiten an, mit 'ihm ins Gespräch zu kommen'" (ZB 2 1995 Anhang: 3).

Als Methoden der formativen Evaluation kamen die teilnehmende, aktive Beobachtung, Fragebögen sowie Schülerinterviews zum Einsatz (ZB 1994: 17ff).

6.3.1.3 *Fortbildungen/Tagungen/modellversuchsbegleitende Transferaktivitäten*

In Zusammenarbeit mit dem Wissenschaftlichen Institut für Schulpraxis ist eine Fortbildungsreihe durchgeführt worden, die parallel zu den Hauptaktivitäten des Modellversuchs organisiert wurde und sich mit "[...] handlungsorientierten Konzeptionen beruflicher Bildung und Ansätze zur Veränderung der an der Verwaltungsschule für die beruflichen Ausbildungsgänge verwendeten Lehrpläne" (ZB 1995: 42) befasst hat.

Darüber hinaus haben eine Projektwoche und ein Aktionstag zu umweltbezogenen Themen in der Verwaltung stattgefunden und es ist ein dienststellenbezogenes Abfallwirtschaftskonzept entwickelt worden (vgl. ZB 1995: 45ff).

6.3.1.4 *Ergebnisse des Modellversuchs aus Sicht der WBL*

Als dokumentierte Ergebnisse auf Modellversuchsebene sind zwei Unterrichtseinheiten und drei Fallstudien veröffentlicht worden. Kernbestandteil des Modellversuchs war die Unterrichtseinheit 'Papier', die nach Ansicht der Mitarbeiter im Modellversuch den am besten 'dokumentierten und reflektierten Teil' der Modellversucharbeit darstellt (vgl. AB 1996: 15). Die zweite Unterrichtseinheit trägt den Titel 'Mobilität' und sei geeignet, exemplarisch Ökologieaspekte im Fach Wirtschaftslehre zu behandeln.

Des Weiteren zeichne sich die Fallstudie 'Beschaffung unter besonderer Berücksichtigung ökologischer Aspekte' durch ein hohes Maß an 'konkreter, berufsbezogener Handlungsorientiertheit' aus. Die anderen beiden Fallstudien seien geeignet, die allgemein-verwaltungsrechtliche Handlungskompetenz in Bezug auf Umweltaspekte, ohne zusätzlichen Stundenaufwand, auszubilden. Hierbei handelt es sich um die Fallstudien 'Entsorgung eines Autowracks' und 'Baumschutz' (vgl. AB 1996: 15f).

6.3.2 *Darstellung des summativen Evaluationsansatzes der WBL im Modellversuch UMWELT*

Die wissenschaftliche Begleitung weist darauf hin, dass der von ihr gewählte 'Begleitforschungsansatz der Aktionsforschung' eine Gewichtung der Einzelinstrumente verlange, "[...] Welche Methode eignet sich für welche Fragestellung? [...]" (ZB 1995 Anhang: 5), damit auch eine 'gebündelte Betrachtung' möglich wird. Ziel sei es, eine möglichst abgesicherte Argumentationspositionen zu erreichen.

In der Dokumentation der Forschungsinstrumente der WBL zu einer Unterrichtseinheit ist auch ein geeigneter Fragebogen zur Bewertung von Texten abgedruckt, der eine Beurteilung der Informationstexte durch die Schüler ermöglichen soll. Gleiches gilt für die Bewertung eines durchgeführten Experiments (vgl. ZB 1995 Anhang: 11).

Weitergehende Angaben zu einer summativen Evaluation oder das Beschreiben von hierfür geeigneten Untersuchungsverfahren unterbleiben.

6.3.3 Ergebnis der Untersuchung - Ausprägungsgrade der Evaluationsstandards

S	N1	N3	N4	N5	N	D1	K1	K5
A	1	1	0	0	0,5	0	0	0

S	K6	K7	K	G1	G2	G3	G4	G5
A	3	0	0,75	1	2	0	0	0

S	G6	G8	G9	G10	G12	G
A	0	0	0	0	0	0,3

S=Standard; A=Ausprägungsgrad

Tab. 6: Ergebnis der metaevaluativen Untersuchung MV Umwelt/Bremen

N1 – Ermittlung der Beteiligten und Betroffenen: Es wird zwischen einer engeren Modellversuchsgruppe und den darüber hinaus gehenden beteiligten Lehrpersonen unterschieden, so dass in diesem Zusammenhang von einer Gruppierung im Hinblick auf Einflussmöglichkeiten gesprochen werden kann. Mitglieder der engeren Modellversuchsgruppe sind drei schulische Vertreter, das Mitglied der behördenbezogenen Projektleitung, der Verantwortliche für den praktischen Teil der Berufsausbildung sowie die beiden Mitarbeiter der wissenschaftlichen Begleitung. Darüber hinaus wird das 'Kooperations-Mitglied' der Modellversuchs-Arbeitsgruppe, die Mitarbeiter des Instituts für Umweltbildung im Beruf (Berlin), genannt (vgl. ZB 1994: 12). Entsprechend der wahrgenommenen Funktionen werden die B&B in die Planungen mit einbezogen. Die betroffenen Schüler werden hinsichtlich ihrer Erwartungen, Wünsche und Interessen befragt, um einerseits die Unterrichtsplanung hieran anzulehnen, andererseits sollen die Schüler Gelegenheit erhalten, ihre Wünsche zu diskutieren (vgl. ZB 1994: 20). Andere Beteiligte und Betroffene, die ein potenzielles Interesse an den Ergebnissen des Modellversuchs haben könnten, wie z. B. Verwaltungsfachschulen anderer Bundesländer oder Großbetriebe, die auch auf Grund ihres Papierverbrauchs u. a. auf die Umwelt Einfluss nehmen, werden nicht ermittelt. Die Ermittlung der B&B bleibt auf die Modellversuchsebene beschränkt, was den Transfer der Ergebnisse erschwert.

N3 – Umfang und Auswahl der Informationen: Die Ergebnisse auf Modellversuchsebene werden ausführlich dargestellt. Dies bezieht sich vor allem auch auf die Dokumentation der erstellten Unterrichtsmaterialien. Für den Leser sind die Berichte leicht unübersichtlich, da der Verlauf des MV nicht zu jeder Zeit nachvollzogen werden kann (ZB 1994, ZB 1995 und AB 1996). In Bezug auf eine summative Evaluation soll ein "[...] guter Fragebogen [...]" (ZB 1995: Anhang 4) eingesetzt werden, der eine Beurteilung eingesetzter Materialien durch die Schüler ermöglichen soll. Die Erkenntnisse hieraus werden jedoch nicht dargestellt. Durch die nicht vollständige Ermittlung der B&B (siehe N1) können auch weitergehende Informationsbedürfnisse nicht ermittelt resp. hervorgehoben werden.

N4 – Feststellung von Werten: Hinsichtlich der grundlegenden Werteorientierung erfolgt im Abschlussbericht der Hinweis auf den Aspekt der ökologischen Orientierung als 'zentrale Säule des Verfassungswesens' (AB 1996: 12). Gleiches gilt für das Verständnis der WBL und der ihr eigenen Aufgaben (ZB 1994: 14). Hier heißt es, dass die Konstruktion und Durchführung von Unterrichtseinheiten im Modellversuch als 'interpretativer Forschungsprozess' verstanden und den 'Forschungsmitteln' ein konstruktives Gewicht zugemessen werde. Die Darstellung der auf Ebene der formativen Evaluation eingesetzten Verfahren ist ausführlich (ZB 1994: 17f). Die Resultate auf MV-Ebene sind direkt auf die Wertmaßstäbe und das eingesetzte Untersuchungsdesign zurückzuführen (ZB 1994: 25f, ZB 1995: 9f, AB 1996: 12f). 'Empirisch gut abgesicherte Rezepte' (vgl. ZB 1994: 14) seien nicht das Ziel der WBL, vielmehr sollten die Erfahrungen problemadäquat sein, um so eine Rekonstruktion und rationale Analyse zu ermöglichen. Wertmaßstäbe für eine summative Evaluation würden daher nicht aufgestellt.

N5 – Klarheit des Berichts: Der zweite Zwischenbericht enthält den Hinweis, dass die Evaluationsziele im Modellversuchsantrag genannt würden. Angaben zu Evaluationsverfahren erfolgen nicht. Die Schlussfolgerungen und Interpretationen beziehen sich auf die Modellversuchsebene.

Resümee Standardbereich N:

Die Transparenz des Vorgehens auf MV-Ebene ist ausreichend, was eine Akzeptanz auf Seiten der unmittelbaren B&B erwarten lässt. Problematisch ist die alleinige Ausrichtung der WBL-Tätigkeiten auf das MV-Feld, was durch das fehlende Ermitteln der weiteren Evaluationsinteressenten (B&B im weiteren Sinne), der Festlegung der Evaluations-Zielformulierung und Werteorientierung u. a. zum Ausdruck kommt. Insgesamt nimmt der Standardbereich der Nützlichkeitsstandards damit ein Niveau ein, das sich im unteren Bereich des ersten Ausprägungsgrades bewegt. Demnach werden die Informationsbedürfnisse der B&B nicht berücksichtigt.

D1 – Praktische Verfahren: Der Standard D1 ist dem Ausprägungsgrad ,0' zuzuordnen, da aus den vorliegenden Berichten (zwei Zwischenberichte und der Abschlussbericht) lediglich ein eingesetztes Evaluationsinstrument (ZB 1995: Anhang 11) zu entnehmen ist, das als 'summatives Evaluationsinstrument' (Fragebogen zur Textbewertung) angesehen werden kann. Die Auswertung dieses Fragebogens ist in den Berichten nicht enthalten.

K1 – Unterstützung der Dienstleistungsorientierung: Aufgabe der Evaluatoren im Sinne dieses Qualitätsstandards ist es unter anderem, über die Interessen von Pädagogen und Organisationen hinaus zu blicken, die Entwicklung der Lernenden zu verbessern und als Ergebnis der Evaluation (Dienstleistungsfunktion), Informationen zu generieren, die für Planungsmodellversuche bzw. bildungspolitische Entscheidungen genutzt werden können. Informationen dieser Art können den Berichten ebenso wenig entnommen werden, wie Aussagen über die Auswirkungen des Modellversuchs im Hinblick auf die Bedürfnisse der B&B (siehe N1).

K5 – Vollständige und faire Einschätzung: Im Rahmen der formativen Evaluation ist eine 'Berufs- und Ausbildungsanalyse' durchgeführt worden, in deren Rahmen auch Ausbilder bremischer Behörden zur Ökologieorientierung ihrer Dienststelle befragt wurden. Der Zeitaufwand für die Beantwortung des Ausbilderfragebogens (vgl. ZB 1994: 42) wird im Nachhinein als zu groß eingeschätzt. Die Aussagefähigkeit der Ergebnisse soll daher nach Angaben der WBL nur 'vorsichtig' wiedergegeben werden. Verbesserungsvorschläge unterbleiben jedoch. Ein ähnliches Vorgehen auf Ebene der summativen Evaluation ist den Berichten nicht zu entnehmen.

K6 – Offenlegung der Ergebnisse: Unter dem Gliederungspunkt 'Kennzeichnungen' werden die Projektleitung und die freigestellten Mitarbeiter im MV, sowie die WBL genannt. Die Berichte sind über den Senator für Bildung und Wissenschaft zu beziehen. Auch die in Auftrag gegebenen Studien und die hierfür Verantwortlichen werden genannt (ZB 1994 und 1995, AB 1996).

K7 – Deklaration von Interessenkonflikten: Interessenkonflikte werden weder dokumentiert, noch werden mögliche Schutzmaßnahmen ex-post vereinbart.

Resümee Standardbereich K:

Der durchschnittliche Ausprägungsgrad der Korrektheitsstandards bewegt sich im unteren Bereich der Ausprägungsgrade, was bedeutet, dass die WBL die gültige Gesetzeslage beachtet, allerdings Werte und Normen der Beteiligten und Betroffenen nicht in ihre Betrachtungen einbezieht. Der positiv abweichende Standard K6 (Offenlegung der Ergebnisse), der die Offenlegung sämtlicher Dokumentationen, Erklärungen u. ä. fordert und im vorliegenden Fall in den höchsten Ausprägungsgrad eingeordnet werden kann, belegt, dass die WBL ihren dokumentarischen Pflichten nachgekommen ist (ZB 1994/ZB 1995/AB 1996), da Ansprechpartner und die Bezugsquellen der Dokumentationen aufgezeigt werden.

G1 – MV-Dokumentation: Es wird hervorgehoben, dass die eigentlichen Untersuchungsschwerpunkte erst im Verlauf des Forschungsprozesses entwickelt werden sollen. Dies geschieht vor dem Hintergrund, dass Fragestellungen veränderbar seien und definitorische Vorüberlegungen auch vollständig aufgegeben werden könnten (ZB 1994: 14). In diesem Zusammenhang wären aber vor allem Änderungen im Verlauf hervorzuheben gewesen (siehe K-Standards), gerade weil die WBL ihr Begleitverständnis als 'konstruktiv' kennzeichnet.

G2 – Kontextanalyse: Das beteiligte Personal sowie der zeitliche Rahmen sind den Zwischenberichten sowie dem Abschlussbericht unter dem Stichwort 'Kennzeichnung' zu entnehmen. Eine Einordnung in den politisch-sozialen Kontext erfolgt insofern, als dass die Besonderheiten der betroffenen Schule beschrieben werden. Des Weiteren kann vermutet werden, da es sich bei dem Auszubildenden um das Land Bremen handelt, dass die Inhalte des MV in die bildungspolitische Zielrichtung eingepasst sind. Weitere entscheidende Faktoren wie parallel laufende Aktivitäten u. a. werden nicht genannt.

G3 – Beschreibung von Zielen und Vorgehen: Dass ein Curriculum für verwaltungs- und bürobezogene Ausbildungsberufe im Land Bremen entwickelt werden soll und welche Evaluationsmethoden auf MV-Ebene verwandt wurden, kann zumindest überblicksartig nachvollzogen werden. Darüber hinaus wird das Vorgehen auf Modellversuchsebene begründet (ZB 1995, Anhang, S. 1).

Die Ziele der Evaluation und das begründete Vorgehen auf dieser Ebene wird jedoch nicht beschrieben bzw. ein summativer Evaluationsansatz wird von der WBL nicht verfolgt (siehe Kap. 6.3.2). Hinweise zur Adaption für andere Nutzer der Modellversuchsergebnisse müssen somit unterbleiben.

Eine Anwendung/Subsumtion der Standards G4, G5, G6, G8, G9, G10 sowie G12 unter die Ausführungen der Modellversuchsdokumentationen kann wegen fehlender Evaluationsaktivitäten auf summativer Ebene nicht erfolgen.

Resümee Standardbereich G:

Eine Auseinandersetzung mit dem Modellversuch fand nur auf Modellversuchsebene statt. D. h. es sind keine Methoden eingesetzt worden, die geeignet sein könnten, die Güte und/oder die Verwendbarkeit des Modellversuchs zu beurteilen und auf diesem Wege für die Fragestellungen, wie sie im Modellversuchsantrag enthalten sein sollen (siehe N5), relevante und ausreichende Informationen hervorzubringen.

Die geringste durchschnittliche Ausprägung der verwendeten Standardbereiche mit mehreren eingesetzten Evaluationsstandards erreicht somit der Standardbereich der Genauigkeitsstandards. Lediglich die Standards G1 (MV-Dokumentationen) und G2 (Kontextanalyse) sind Ausprägungsgraden zuzuordnen, die aus dem schwächsten Ausprägungsgrad herausragen. Dies ist für den Standard G2 damit zu begründen, dass die Besonderheiten der 'Verwaltungsschule' im politisch-sozialen Kontext des Landes Bremen hervorgehoben werden (ZB 1994: 7). Im Bereich des Standards G1 wird auf den Evaluationsansatz verwiesen und hervorgehoben, dass der Untersuchungsgegenstand im Verlauf des MV entwickelt werden soll (ZB 1994: 14). Diese Entwicklung wird allerdings nicht dargestellt.

Die Anwendung des Kongruenzschemas zur Feststellung der gegenseitigen Abhängigkeit von Modellversuchsaufbau inkl. Evaluationsverfahren und Verfahren der Zielermittlung konnte nicht durchgeführt werden, da die notwendigen Ziele der Evaluation (Prüfaspekt 1) nicht vorhanden sind.

Zwischenergebnis – Analyse und Interpretation der Ergebnisse:

Durch die schwache Ausprägung der Nützlichkeitsstandards wird die nicht ausreichende Auseinandersetzung der Evaluatoren mit Beteiligten und Betroffenen deutlich. Ihre Informationsbedürfnisse wurden nicht in der Art ermittelt, als dass hierauf aufbauend eine zielgruppenadäquate Evaluation hätte entwickelt werden können.

Dieses Problem wirkt sich auch auf die Ausprägung der Korrektheitsstandards aus, da so nicht die Interessen und Bedürfnisse des gesamten Zielgruppenspektrums berücksichtigt werden können. Mit Ausnahme des Standards K6 (Offenlegung der Ergebnisse), der sich im Wesentlichen auf die Dokumentationspflicht, die auch

durch die BLK gegeben ist, bezieht, sind alle anderen Standards nicht erfüllt worden. Zwar hebt die WBL hervor, dass ihre Tätigkeiten nicht vom Gesamtprozess isoliert betrachtet werden können, sondern als 'integraler Aspekt' des Gesamtprojektes anzusehen ist und Fragestellungen sich verändern oder auch aufgegeben werden können (vgl. ZB 1994: 14f), dennoch sind keine Aussagen über den geplanten Verlauf und mögliche Änderungen in den Dokumentationen enthalten. Hieraus folgt, dass das Recht auf 'Beeinträchtigungsfreiheit' der Beteiligten und Betroffenen nicht ausreichend in die Betrachtungen einbezogen wird; gleiches gilt für die Rechte anderer in Bezug auf umfassende Information.

Mit Ausnahme der Standards G1 (MV-Dokumentationen) und G2 (Kontextanalyse) konnte keiner der Genauigkeitsstandards erfüllt werden, da die WBL keine summarische Evaluation durchgeführt hat (Korrelation mit Standard D1).

6.4 Darstellung und Evaluation des Modellversuchs SCHLÜSSELQUALIFIKATIONEN

6.4.1 Beschreibung des Modellversuchs

6.4.1.1 Allgemeine und strukturelle Angaben

Der Modellversuch 'Schlüsselqualifikationen für Kaufleute im Einzelhandel' ist im Bundesland Hamburg in der Zeit vom 01.01.1991 bis 31.12.1993 durchgeführt worden. Die Projektleitung hatte der Oberschulrat des Amtes für Schule der Hansestadt Hamburg übernommen.

In den Modellversuch einbezogen waren vier Hamburger Schulen mit insgesamt neun Klassen, in denen Einzelhandelskaufleute in unterschiedlichen Branchen unterrichtet wurden. Die Bandbreite des Sortiments reichte von Lebensmitteln über Computer/Audio/Video, Heimwerkerbedarf, Hausrat und Juwelierartikel, Textil bis zu Kraftfahrzeugen (vgl. ZB 1992: 6).

6.4.1.2 Modellversuchsziele und -inhalte/Tätigkeiten der WBL auf Modellversuchsebene

Aufgrund der geänderten Qualifikationsanforderungen, die auch ihren Ausdruck in dem 1987 neu geordneten Rahmenlehrplan finden und sich in einer stärker formulierten 'allgemeinen fach- und berufsübergreifenden Qualifizierung' ausdrücken, wurde das Konzept der Schlüsselqualifikationen verstärkt diskutiert (vgl. ZB 1992: 4f). In diesem Zusammenhang ist auch der Modellversuch initiiert worden, der folgender Formal-Zielsetzung nachgeht:

"Erprobung und Entwicklung unterschiedlicher didaktischer Formen der Situations- und Handlungsorientierung, um ihre Einsatzmöglichkeiten einzuschätzen und verschiedene Formen der Vermittlung von Schlüsselqualifikationen verfügbar zu machen" (ZB 1992: 5).

Dabei wurde ein Verständnis des Begriffes 'Schlüsselqualifikationen' zu Grunde gelegt, das seinen Ausdruck in den Bereichen 'persönlich-charakterliche Grundfä-

higkeiten, kommunikative Fähigkeiten und allgemeine kognitive Leistungsfähigkeit' fand (vgl. ZB 1992: 12).

Die zu entwickelnden und zu erprobenden Organisationsformen wirtschaftsberuflichen Lernens, auch Modellversuchsvarianten genannt, strebten "[...] einerseits einen hohen Grad von Annäherung an berufliche Realität [...] und andererseits durch besondere didaktische Maßnahmen angemessene Komplexität und Ganzheitlichkeit des Lernhandelns [...]" (ZB 1992: 16) an.

Wesentlicher Bestandteil aller (drei) Modellversuchsvarianten war die Tatsache, dass sie Modelle von Einzelhandelsunternehmen zur didaktischen Grundlage hatten und sich durch den ihnen eigenen innovativen Charakter von den herkömmlichen Vermittlungsarten unterscheiden sollten.

Die erste Modellversuchsvariante 'Unternehmensprojekte' stellte ein reales Einzelhandelsunternehmen mit realen Ein- und Verkaufsinteraktionen, realer Übertragung dieser Vorgänge in das Rechnungswesen usw. dar.

Trotz der Reduktion der Modellversuchskonstruktion wurde festgestellt: "Die Konzipierung des Einzelhandelsunternehmens als Projekt bewirkt einen relativ hohen Grad an Komplexität. Insbesondere durch Überschreiten der Schulgrenze und Einbeziehung des realen Marktes in Ein- und Verkaufsaktionen kommt es zu authentischen Handlungen, die eine Identität von Lern- und Arbeitshandeln erlauben" (ZB 1992: 19).

Die Modellversuchsvariante 'Lernbetrieb' war als Simulation eines Unternehmens konzipiert, so dass die Auszubildenden simulativ handelnd an Geschäftsfällen teilnehmen konnten, die typisch für einen Einzelhandelsbetrieb sind.

'Strategie' als dritte zu erprobende Organisationsform stellte mehrere Einzelhandelsunternehmen dar, die in einer Konkurrenzsituation zueinander standen; die Außenwelt wurde EDV-gestützt repräsentiert. Ein Schwerpunkt in diesem Modell war die Betonung der Förderung von Entscheidungsfähigkeit, Verhandlungsfähigkeit und Selbständigkeit in einem stark prozessbetonten Lernsystem. Regulärer Unterricht wurde als 'Serviceleistung' integriert (vgl. ZB 1992: 19ff).

Der Grundgedanke bei der Konstruktion der Modelle war die didaktische Integration eines oder mehrerer Modelle von Einzelhandelsunternehmen in den Unterricht, um eine Lernorganisation zu erreichen, bei der die Systeme Unterricht und Unternehmensmodell eine Symbiose eingehen, um durch die Kombination die Funktionstüchtigkeit des daraus entstehenden Gesamtsystems qualitativ steigern zu können (vgl. ZB 1992: 16f).

Die wissenschaftliche Begleitung erfolgte durch den Fachbereich Erziehungswissenschaft der Universität Hamburg. Aus dem Abstimmungsprozess von Bildungspraktikern und Wissenschaftlern ist vor Beginn des Modellversuchs eine Option für ein Begleit- und Evaluationskonzept entstanden, "[...] das dem Muster der Steuerungs- und Optimierungsfunktion folgt (Helfer- und Beratungsmodell)" (ZB 1992: 23).

Diese Art des Aufgabenverständnisses der WBL präjudiziert die Aktions- und Handlungsforschung als konzeptionelle Grundlage der Modellversuchsbegleitung. Darüber hinaus war zu berücksichtigen, dass Modellversuche über einen längeren Zeitraum nicht invariabel sind, so dass nach Ansicht der Mitglieder der WBL die Handlungsforschung durch eine ‚vorsichtige Methodenerweiterung‘ verändert werden musste, damit die gewonnenen Erkenntnisse durch ‚intersubjektive Überprüfbarkeit‘ gefestigt werden konnten (vgl. ZB 1992: 23).

Die Mitglieder der WBL hoben drei Voraussetzungen hervor, die für den gewählten Ansatz von Bedeutung waren:

"Erstens wird sie (die Aktions- und Handlungsforschung; Anm. Verf.) von einer zu Beginn zu formulierenden Hypothese geleitet, derzufolge es gilt, einen Bereich der Erziehungswirklichkeit umzugestalten. Zweitens ist es erforderlich, dass alle Beteiligten, die Theoretiker und die Praktiker, ihre Arbeit aus einer gemeinsamen wissenschaftsgeleiteten Verantwortung heraus gestalten. Ein drittes Merkmal dieser Forschung ist, dass die Teilergebnisse und Einsichten jeweils wieder als Korrektiv und Impuls dem kontinuierlichen Prozess des Modellversuchs eingegeben werden" (ZB 1992: 23). Letzteres macht deutlich, dass die Evaluation als Prozess mit Rückkoppelung zum Modellversuchsverlauf angelegt ist und somit einer formativen Evaluation entspricht.

6.4.1.3 Fortbildungen/Tagungen/modellversuchsbegleitende Transferaktivitäten

Zu Beginn des Modellversuchs hat ein Halbtagesseminar zum Thema 'Qualifikationsveränderungen im Einzelhandel' stattgefunden, um beteiligte Lehrerinnen und Lehrer für den Themenkomplex zu sensibilisieren. Darüber hinaus wurden zwei Fachtagungen zu den Themen 'Qualifikation durch Innovation' (mehrtägig) und 'Schlüsselqualifikationen für Kaufleute im Einzelhandel' (eintägig) durchgeführt (vgl. ZB 1993: 63ff). Außerdem ist ein Berichtband mit dem Thema 'Schlüsselqualifikationen im Einzelhandel' erschienen und es sind sechs Diplomarbeiten zu Themenbereichen aus dem Modellversuch angefertigt worden (vgl. AB 1994: 26f).

6.4.1.4 Ergebnisse des Modellversuchs aus Sicht der WBL

- Die Schlüsselqualifikation 'Denken in Zusammenhängen' wurde nach Ansicht der WBL in der Modellversuchsvariante 'Unternehmensprojekt' durch die Konfrontation mit Rückmeldungen von Kunden, Lieferanten oder Konkurrenten gefördert. Im Hinblick auf die MV-Variante 'Lernbetrieb' sei es die Simulation zweier Betriebe, in die fächerübergreifende komplexe Aufgabenstellungen integriert wurden und die auf diesem Wege zu einer Förderung dieser Schlüsselqualifikation beigetragen haben. Dies belege auch der Einsatz des sog. 'Beziehungsfensters', das vor und nach Durchführung des Modellversuchs die Ausprägung der Schlüsselqualifikation geprüft habe. In der Modellversuchsvariante 'Strategie' dagegen wurde durch die Konkurrenzssituation zu anderen Unternehmen und durch den Prozesscharakter der Simulation vernetztes Denken gefördert, da relevante Fachinhalte als 'Handlungsstrang' in ein Problem integriert wurde (vgl. AB 1994: 27ff).

- Ferner kommt die WBL in Bezug auf die Förderung der Schlüsselqualifikation 'Problemlösefähigkeit' im Rahmen der Ergebnisse der Modellversuchsvariante 'Unternehmensprojekt' zu dem Schluss: "Versuch und Irrtumsprozesse mit realem Feedback führen über Diskrepanzerlebnisse zur Förderung der Problemlösefähigkeit" (vgl. AB 1994: 27). Darüber hinaus hätten die Schüler in der MV-Variante 'Strategie' versucht, Probleme grundsätzlich allein anzugehen, indem sie die zur Lösung notwendigen Strategien selbständig entwickelt hätten (vgl. AB 1994: 32). Eigeninitiative und Selbstverantwortungsbereitschaft resultierten außerdem aus der Identifikation mit dem Ladenprojekt (Modellversuchsvariante 'Unternehmensprojekt'; vgl. AB 1994: 27).
- Kooperations- und Teamfähigkeit als Elemente einer persönlichen und beruflichen Handlungsfähigkeit hätten sich nach Ansicht der WBL aus der Notwendigkeit der Abstimmung mit den Interessen anderer Gruppenteilnehmer vor allem in der MV-Variante 'Unternehmensprojekt' ausgebildet. Aber auch die starke Betonung der Gruppenarbeit als Sozialform in der Modellversuchsvariante 'Lernbetrieb' habe die Entwicklung der Kooperationsfähigkeit gefördert. Hohe Anforderungen im Bereich der sozialen Kompetenzen seien außerdem im Projekt 'Strategie' vorhanden gewesen, so dass die Kooperations- und Konfliktfähigkeit durch die enge Zusammenarbeit der Teammitglieder bei der Bewältigung von Problemen unterstützt werden konnte (vgl. AB 1994: 27ff).
- Neben den schon genannten Schlüsselqualifikationen seien auch kommunikative Fähigkeiten, darunter auch die Präsentationsfähigkeit, in der Modellversuchsvariante 'Strategie' ausgebildet worden (vgl. AB 1994: 32).

Generalisierend, d. h. über die entworfenen 'Modellversuchsvarianten' hinaus, hat die WBL Empfehlungen ausgesprochen, die sich auf eine Veränderung schulischer Lernplätze in Richtung authentischer, simulierender und/oder symbolischer Arbeitshandlungen, der Sicherung kognitiver Komplexität durch fächerübergreifende und transferfördernde Lernprozesse und durch eine veränderte Lehrerrolle (Lehrer als Organisator von Lernprozessen und Lernberater) beziehen, um zu einer Förderung von Schlüsselqualifikationen durch schulische Lernprozesse beizutragen (vgl. ZB 1992: 14f).

6.4.2 Darstellung des summativen Evaluationsansatzes der WBL im Modellversuch SCHLÜSSELQUALIFIKATIONEN

Die 'allgemeine' Evaluationsaufgabe bestehe darin, "[...] festzustellen, ob die mit dem Modellversuch gegebenen lernorganisatorischen Bedingungen in der Lage sind, die angestrebte Zielerreichung, nämlich die Schlüsselqualifikationen, zuzulassen bzw. zu fördern" (ZB 1992: 24).

Daher müsse über die Bemühungen der formativen Evaluation hinaus, wenn es um die Förderung von Schlüsselqualifikationen durch die Lernorganisation gehe, auch der Lernfortschritt ermittelt werden, der durch die Lernorganisation ermöglicht wird. Aus diesem Grund seien in die einzelnen Modellversuchsvarianten Instrumen-

te integriert, die eine Eingangsevaluation ermöglichen. Spätere Zwischen- bzw. Abschlussevaluationen sollen Einsicht in die Lernzuwächse in Bezug auf die Schlüsselqualifikationen geben. Jeweilige Kontrollgruppen dienen der Feststellung der Qualifikationsdifferenz zwischen Modellversuchsschülern und solchen, die nicht am Modellversuch beteiligt sind (vgl. ZB 1992: 24). Hierbei wird darauf geachtet, dass Klassen mit Schülern vergleichbarer Schulbildung ausgewählt werden.

Zu Beginn der Ausbildung zum Einzelhandelskaufmann wurden die am Modellversuch beteiligten Auszubildenden mit einem Verkaufsprojekt konfrontiert, um die Eingangs-Schlüsselqualifikationen zu ermitteln. Jede am Modellversuch beteiligte Klasse verkaufte an zwei Tagen innerhalb des Schulgebäudes Waren, wobei hinsichtlich des anzubietenden Sortiments, der Warenpräsentation und des Standortes innerhalb des Gebäudes freie Wahl bestand.

Die Auswertung sollte mit Hilfe eines Beobachtungsbogens erfolgen. Auf Grund von Schwierigkeiten bei der Einschätzung des Gesehenen durch Lehrer ist die WBL zu der Überzeugung gelangt, dass das Verkaufsprojekt im Hinblick auf seinen Zweck nicht geeignet sei. Daher wurde an Stelle der Verkaufssituation ein Rollenspiel eingesetzt, für das warenbezogene Situationen vorgegeben waren und das zeitlich begrenzte Verkaufsgespräche darstellte. Die Dokumentation erfolgte durch Videoaufnahmen; die Auswertung dieser Aufnahmen dagegen mittels eines Rasters, das den vorhandenen Wortschatz, die Artikulation, Argumentation und Körpersprache einschätzen ließ (vgl. ZB 1992: 33).

Darüber hinaus mussten alle Schüler (auch die der Kontrollgruppe) eine berufsbezogene 'komplexe Ernstsituation' in Form einer Falldarstellung schriftlich bewältigen. Die Auswertung erfolgte durch ein von Reetz entwickeltes Schema, mit dem den 'zielrelevanten Aktivitäten' Punkte zugeordnet wurden. Beide Messinstrumente sollten zu Beginn und nach Abschluss der Unternehmensprojekte zum Einsatz kommen (vgl. ZB 1992: 32).

In der Modellversuchsvariante 'Unternehmensprojekte' sind die für die zu erwerbenden Schlüsselqualifikationen vorhandenen Fähigkeiten über Fragebogen (SQ - Initiative und Verantwortung), Situationsaufgaben (SQ – Problemlösung und analytisches Denken) und Rollenspiele (SQ – Verhandlungsfähigkeit) ermittelt worden.

Der erste Teil des Fragebogens bezog sich auf die 'Persönlichkeitsforschung im Verhaltensbereich', der zweite Teil dagegen auf 'Situationsforschung'. Der private und berufliche Bereich wurde gleich stark gewichtet, um Aussagen für beide Bereiche erhalten zu können. Der Einsatz des Fragebogens erfolgte gleichzeitig in der Experimental- und der Kontrollgruppe, wobei eine erste Befragung zu Beginn ihrer Ausbildung und eine zweite nach Durchlaufen des Unternehmensprojektes durch die Experimentalgruppe durchgeführt wurde. Die Auswertung der Situationsaufgaben erfolgte ebenso wie die Eingangs-Fallsituation mit dem von Reetz entwickelten

Schema (Zuordnung von Punkten zu zielrelevanten Antworten). Zur Auswertung der Rollenspiele s. o. (vgl. ZB 1992: 33).

Im 'Lernbetrieb' als zweite Modellversuchsvariante sind ein sog. 'Beziehungsfenster' (SQ – vernetztes Denken), ein Aufgabenset (SQ - Problemlösung und analytisches Denken) und Videoaufzeichnungen, die durch ein Auswertungsschema im Ratingverfahren ausgewertet wurden (SQ – kommunikative Fähigkeiten) zum Einsatz gekommen.

Im Beziehungsfenster mussten Begriffe (u. a. kaufmännische Grundbegriffe) in Relation zueinander gesetzt und die Art der Beziehung benannt werden. Für die Auswertung wurden jeweils die vom Begriff wegführenden Pfeile gewertet, sofern sie von der WBL als relevant (Abstufung vorhanden) anerkannt wurden. Die unterschiedliche Qualität der Beziehungsangaben wurde außerdem gewichtet (vgl. ZB 1992: 75). Zur Auswertung des Aufgabensets und der Videoaufnahmen siehe oben.

Im Rahmen der Modellversuchsvariante 'Strategie' kamen keine Methoden der summativen Evaluation zur Wirkung, da diese Variante in besonderem Maße von den 'Lernzielen' bzw. den eingenommenen 'lerntheoretischen Positionen' der Unterrichtsgestalter abhängt. Eine summative Evaluation könne nur Erkenntnisse zum Zeitpunkt 't' hervorbringen und würde somit nicht das dynamische System dieser speziellen Unterrichtsorganisation erfassen lassen (vgl. ZB 1992: 24ff).

6.4.3 Ergebnis der Untersuchung - Ausprägungsgrade der Evaluationsstandards:

S	N1	N3	N4	N5	N	D1	K1	K5
A	1	0	3	2	1,5	3	1	0

S	K6	K7	K	G1	G2	G3	G4	G5
A	3	0	1,0	1	2	2	2	1

S	G6	G8	G9	G10	G12	G
A	1	1	0	0	0	1,0

S=Standard; A=Ausprägungsgrad

Tab. 7: Ergebnis der metaevaluativen Untersuchung MV SQ/Hamburg

N1 – Ermittlung der Beteiligten und Betroffenen: Die in den Modellversuch eingebundenen Klassen werden hinsichtlich Klassenbezeichnung und Branchenschwerpunkt genannt. Über die Auswahl der Klassen sowie der beteiligten Lehrer enthält der Bericht keine Informationen. Weitere Angaben zum Informationsweg und dem Austausch über Bedürfnisse der Beteiligten fehlen. Lediglich die Namen der involvierten Lehrer und der Mitglieder der wissenschaftlichen Begleitung werden genannt (vgl. ZB 1992: 6). Andere potenzielle Beteiligte und Betroffene werden ebenfalls nicht ermittelt, obwohl die Förderung und Messung von Schlüsselqualifikationen (Zuwachs) berufsübergreifend von Interesse ist.

N3 – Umfang und Auswahl der Informationen: Zwischen- und Abschlussberichte enthalten u. a. allgemeine Ausführungen zum Konzept der Schlüsselqualifikationen sowie dem arbeitsmarktpolitischen und betriebswirtschaftlichen Hintergrund dieses Ansatzes (vgl. ZB 1992: 13f, 21f, AB 1994: 12f). Gleiches gilt für die Funktion von Evaluation (vgl. ZB 1992: 22f). Darüber hinaus weisen die Berichte Informationen im Hinblick auf die Einordnung des Schlüsselqualifikationskonzeptes in den handlungsorientierten Ansatz auf (vgl. ZB 1992: 13f, AB 1994: 12f), was zu einer Überfrachtung der Dokumentationen führt. Außerdem sind die Berichte eher unübersichtlich gestaltet (z. B. fehlerhafte Nummerierung der Anlagen), so dass die Gefahr besteht, relevante Angaben zu übersehen; Informationen und Ergebnisse können nur mit großer Mühe entnommen werden. Generell ist durch die wenig ausreichende Ermittlung der B&B (Standard N1) nicht gewährleistet, dass die relevanten Informationen selektiert wurden.

N4 – Feststellung von Werten: Die der Modellversuchsevaluation zu Grunde liegenden Wertmaßstäbe lassen sich einerseits aus dem verfolgten Konzept der Schlüsselqualifikationen implizit entnehmen, andererseits aus der die Untersuchung leitenden Eingangshypothese erkennen (vgl. ZB 1992: 23, AB 1994: 26). Hinsichtlich der Evaluationsstätigkeiten auf der Objektebene (formative Evaluation) hebt die WBL hervor, dass sie sich in Auseinandersetzung mit den Bildungspraktikern für ein Konzept nach dem Muster der Steuerungs- und Optimierungsfunktion (Handlungs- und Aktionsforschung) entschieden hat (vgl. ZB 1992: 23). Im Hinblick auf die summative Evaluation hat sich die WBL für 'Evaluationsansätze' entschieden, die sich "[...] auf der Linie der Intentionen einer empirisch erweiterten Handlungsforschung [...]" bewegen (ZB 1992: 23). Es ist somit erkennbar, dass keine einseitige Verfechtung des qualitativen oder quantitativen Paradigmas verfolgt wird, sondern dass durch die Kombination von Aspekten/Methoden beider Ausrichtungen eine Optimierung des eigenen Evaluationsansatzes durch Minimierung der Nachteile bei einseitiger Evaluationsgestaltung vermieden werden soll (vgl. Kap. 4).

N5 – Klarheit des Berichts: Der erste Zwischenbericht (ZB 1992: 24) enthält die Formulierung der Evaluationsaufgabe, die auch als Zielbestimmung angesehen werden kann. Eine Operationalisierung unterbleibt jedoch, so dass die durchgeführte Eingangsevaluation und Ergebnisfeststellung (vgl. AB 1994: 26) nicht aussagefähig sein kann. Die für den Einsatz geplanten Evaluationsinstrumente werden skizziert.

Die Konsequenzen und Empfehlungen, die aus den Untersuchungen heraus vorgeschlagen werden, stellen lediglich Thesen dar (vgl. AB 1994: 32f). Durch die fehlende Einordnung in den Begründungszusammenhang sind die Thesen nicht geeignet, 'handlungsanleitend' den 'Transfer-, Implementationsvorgang' zu initiieren. Die Aussagen haben lediglich Aufforderungscharakter.

Resümee Standardbereich N:

Da die Beteiligten und Betroffenen hinsichtlich ihrer Informationsbedürfnisse nicht ausreichend ermittelt werden, können deren Interessen nicht konsequent berücksichtigt werden.

Der Ausprägungsgrad '1' des Standards N1 (Ermittlung der B&B) ergibt sich aus der Benennung der beteiligten Schulen und der Anzahl der Klassen (vgl. ZB 1992: 6), weitergehende Angaben zum Informationsweg und dem Austausch der betroffenen Schulpraktiker erfolgen nicht. Der Standard N3 (Umfang und Auswahl der Informationen) wird hinsichtlich seiner Ausprägung dem Grad '0' zugeordnet, da die Dokumentationen zum Teil undurchsichtig sind. Dies ist u. a. mit der 'übermäßig' ausführlichen Darstellung des Konzepts der Schlüsselqualifikationen sowie dem arbeitsmarktpolitischen und betriebswirtschaftlichen Hintergrund hierzu zu begründen (vgl. ZB 1992: 13f, 21f u. AB 1994: 12f). Das Erreichen des höchsten Ausprägungsgrades durch den Standard N4 (Feststellung von Werten) kann mit den aus dem Konzept der Schlüsselqualifikationen zu entnehmenden Wertmaßstäben und der leitenden Eingangshypothese begründet werden (vgl. ZB 1992: 23f, AB 1994: 26). Zu der 'Klarheit des Berichts' (N5) trägt die Formulierung der Evaluationsaufgabe bei, allerdings unterbleibt deren Operationalisierung (vgl. AB 1994: 26). Des Weiteren werden die Konsequenzen und Empfehlungen nur in Form von Thesen formuliert, umsetzbare Handlungsempfehlungen fehlen, so dass die Nützlichkeitsstandards insgesamt nur ein Standard-Niveau erreichen, das sich im unteren Drittel der Ausprägungsgrade bewegt.

D1 – Praktische Verfahren: Die WBL hat sich für ein Evaluationskonzept auf Objektebene im Sinne der Aktions- und Handlungsforschung entschieden (ZB 1992: 23). Das zu entwickelnde Lernsystem, bestehend aus sieben verschiedenen Varianten, hat die Evaluatoren nach eigenen Angaben überfordert (ZB 1993: 52), da die personelle Ausstattung nicht gegeben war. Hieraus lässt sich schließen, dass wegen der unzureichenden Kontakthäufigkeit ein reibungsloser Verlauf des MV nicht mehr gewährleistet war und Störungen eingetreten sind. Auf eine kritische Phase dieser Art wird auch von den Evaluatoren verwiesen (ZB 1993: 52). Dennoch konnte die Evaluation zu Ende geführt werden, da Studenten in die Evaluationstätigkeiten einbezogen wurden.

Im Hinblick auf die summativen Evaluationstätigkeiten zur Feststellung von Lernzuwächsen ist das Initiieren von ex-ante- und ex-post-Erhebungen angemessen. Die WBL hat für die unterschiedlichen Kompetenzschwerpunkte innerhalb der Bandbreite der Schlüsselqualifikationen in den Modellversuchsvarianten ein vielschichtiges Instrumentarium entwickelt, um der Heterogenität der Modellversuchsvarianten Rechnung zu tragen. Die Abstimmung des notwendigen Evaluationsinstruments mit der zu bewertenden Schlüsselqualifikationsausprägung ist sinnvoll vorgenommen worden:

- Denken in Zusammenhängen – Beziehungsfenster: Ein aus der pädagogischen Psychologie stammendes Verfahren, das auch schon in anderen Modellversuchen zur Evaluation von Planspielen (MV WOKI) mit dem Ziel zum Einsatz gekommen ist, Beziehungen zwischen betriebswirtschaftlichen Begriffen herzustellen.

- Problemlösefähigkeit – Fallstudie: Die eingesetzte Fallstudie zielt darauf ab, dass die Schüler kausales betriebswirtschaftliches Denken ebenso demonstrieren wie auch konstruktiv problemlösendes Denken.
- Kommunikative Fähigkeiten – Kurzvortrag auf Grundlage eines vorgegebenen Fragerasters: Die mit Hilfe von Fragen notierten Stichworte müssen vor einer Gruppe (Videoaufzeichnung) zu einem zusammenhängenden Vortrag ausformuliert werden.
- Persönlich charakterliche Grundfähigkeiten – Befragungen: Mit Hilfe eines Fragebogens wird der Zuwachs an Initiative und Verantwortung gemessen.
- Allgemeine kognitive Leistungsfähigkeit und kommunikative Fähigkeiten - Eingangsprojekte/Rollenspiele: Schülerleistungen in Verkaufsphasen werden auf Grundlage von Beobachtungen, die mit Hilfe von Videoaufnahmen dokumentiert sind, mittels eines einheitlichen Auswertungsraster bewertet.

Die ausgewählten Instrumente führen nach Lage der Berichte zu einem vertretbaren Maß an 'Störungen', im Sinne von Unterbrechungen, auf der Objektebene.

K1 – Unterstützung der Dienstleistungsorientierung: Die inhaltliche Ausrichtung des Modellversuchs warf als Ergebnis eines Evaluationsschrittes das Problem mangelnder Akzeptanz der Projekte durch die Schüler auf. Diese wiesen auf den Charakter der Abschlussprüfungen hin, die rezeptives Wissen und nicht prozedurales Wissen, wie es durch die Modellversuchsvarianten vermittelt werde, abfragen (vgl. ZB 1993: 54). Es erfolgt daraufhin lediglich der Hinweis im Abschlussbericht, dass es erforderlich sei, dass im dritten Jahr der Ausbildung zur herkömmlichen Struktur des Unterrichts zurück zu kehren ist (vgl. AB 1994:90). Das Auftreten dieser Kritik ist u. a. auf Mängel bei der Einhaltung des Standard N1 (Ermittlung der Beteiligten & Betroffenen) zurückzuführen.

Positiv ist, dass der Bericht der WBL zu den Evaluationstätigkeiten dieses Problem ausweist, jedoch wird es nicht reflektiert, d. h. auf seine Ursachen zurückgeführt (Ausgestaltung des MV und nicht ausreichende Ermittlung der B&B – Standard N1). Aufgabe der Evaluatoren im Sinne dieses Qualitätsstandards ist es jedoch, über die Interessen von Pädagogen und Organisationen hinaus zu blicken, die Entwicklung der Lernenden zu verbessern und als Ergebnis der Evaluation (Dienstleistungsfunktion) Informationen zu generieren, die für Planungsmodellversuche bzw. bildungspolitische Entscheidungen genutzt werden können. Nach Lage der Berichte sind die Erkenntnisse demnach nur bedingt geeignet, die Anforderungen des Standards zu erfüllen.

K5 – Vollständige und faire Einschätzung: Auf Stärken und Schwächen des Modellversuchs wird generell nicht eingegangen. Das Problem der Diskrepanz zwischen Modellversuchsausrichtung und Gestaltung der Abschlussprüfung (siehe K1) wird nicht als Problem erkannt, es werden demnach nur die Modellversuchsvarianten dokumentiert, d. h. beabsichtigte Erkenntnisse hervorgehoben.

K6 – Offenlegung der Ergebnisse: Die Zwischen- und Abschlussberichte sind über das Amt für Schule (Hamburg) zu beziehen. Hierin wird ein Projektleiter genannt, der auch als Ansprechpartner fungiert. Darüber hinaus enthält der Abschlussbericht Hinweise, in welchen Veröffentlichungen die Befunde zur Evaluation außerdem noch dokumentiert wurden (vgl. AB 1994: 26f).

K7 – Deklaration von Interessenkonflikten: Es werden keine Ursachen für mögliche Interessenkonflikte der beteiligten Parteien genannt. Entsprechend fehlen schriftliche Vereinbarungen über Schutzmaßnahmen hierzu.

Resümee Standardbereich K:

Es ist deutlich geworden, dass die Interessen der B&B nicht ermittelt wurden. Diese Problematik zeigt sich vor allem in der Ausprägung des Standards K1: Hier hat man das Problem der Organisation und Ausrichtung der Abschlussprüfungen vernachlässigt, was sich auf die Qualität der Evaluation im Sinne ihrer 'Dienstleistungsfunktion' auswirkt. Obwohl die betroffenen Schüler Schwierigkeiten bei der Akzeptanz der Projekte hervorhoben (Standard K5), erkannten die Evaluatoren hierin kein Problem auf der Objektebene und unterbreiteten somit auch keine Vorschläge für die Vermeidung dieses Aspektes in einem Folge-Modellversuch, was die Qualität der Evaluation als Beitrag zur 'bildungspolitischen Entscheidungshilfe' beeinträchtigt. Den Ausprägungsgrad '3' erreicht der Standard K6 (Offenlegung der Ergebnisse; AB 1994: 26f), da die Bezugsquellen für die Modellversuchsdokumentationen ebenso genannt werden wie Ansprechpartner und weitere Veröffentlichungen. Die Deklaration von Interessenkonflikten der Evaluatoren resp. ex-ante-Maßnahmen für ihre Vermeidung (Standard K7) wurde nicht unternommen. Insgesamt erreicht dieser Standardbereich nur ein Qualitätsniveau, das in etwa der ersten Ausprägungsstufe zuzuordnen ist, so dass die Aussage gerechtfertigt scheint, dass die Evaluatoren die Werte und Normen der B&B nicht bzw. nur sehr begrenzt berücksichtigen.

G1 – MV-Dokumentation: Obwohl die WBL ihre Tätigkeiten auf der Modellversuchsebene nach dem 'Helfer- und Beratermodell' ausrichten will und die Bedeutung der formativen Evaluation hervorhebt, werden Änderungen im Modellversuchsprozess nicht dokumentiert. Man stellt zwar heraus, dass gerade die Veränderlichkeit des Modellversuchsprozesses ein Merkmal von Modellversuchen sei (vgl. ZB 1992: 23) und dass diese Veränderungen geradezu erwünscht seien, dennoch werden alle Modellversuchsvarianten nur als 'statische' Produkte dargestellt.

G2 – Kontextanalyse: Es erfolgt eine Charakterisierung des Konzepts der Schlüsselqualifikationen, das den Tätigkeits- und Qualifikationsanforderungen des 'Dienstleistungsbereiches Einzelhandel' gegenüber gestellt wird: Aufgrund des (Personal-) Einsatzes in vielfältigen Vertriebsformen, der unterschiedlichen Warenpräsentationen und Warenkommunikation sei hier eine flexible berufliche Handlungsfähigkeit notwendig, die durch eine entsprechende Organisation des Unterrichts ausgebildet werden soll (vgl. ZB 1992:10ff). Darüber hinaus erfolgt eine zeitliche und geographische Verortung des Modellversuchs. Das politische oder soziale

Umfeld bzw. parallel laufende konkurrierende Aktivitäten werden jedoch nicht dargestellt.

G3 – Beschreibung von Zielen und Vorgehen: Der erste Zwischenbericht enthält eine Eingangshypothese und die Beschreibung der allgemeinen Evaluationsaufgabe. Eine Operationalisierung mit genauer Zielformulierung unterbleibt, so dass lediglich Formalziele vorliegen (vgl. ZB 1992: 23f). Das Vorgehen bei der Evaluation, d. h. das Durchführen einer Eingangs- und späteren Zwischen- und Endevaluation sowie die zum Einsatz kommenden Instrumente werden erläutert und begründet (vgl. ZB 1992: 23ff, 32ff; AB 1994: 25ff; siehe auch Kap. 6.3.2 und Standard D1). Wegen der fehlenden Operationalisierung, die das Ausmaß der Zielerreichung nicht einschätzen lässt, muss die Einordnung in den höchsten Ausprägungsgrad unterbleiben.

G4 – Verlässliche Informationsquellen: Die Evaluatoren dokumentieren begründend den Einsatz der verschiedenen Evaluationsinstrumente: Es wird deutlich, welche Schlüsselqualifikationen in der jeweiligen Modellversuchsvariante im Besonderen ausgebildet werden und mit welchen Instrumenten/Methoden das erreichte Ausmaß der Kompetenzen ermittelt werden soll. Hierdurch kann die Stärke der Beeinflussung der Ergebnisse durch die Probanden zumindest bedingt, d. h. in Bezug auf die Einflussmöglichkeit des Probanden auf das Ergebnis der Datenerhebung im Rahmen der jeweiligen Evaluationsmethode, eingeschätzt werden. Offen bleibt jedoch, nach welchen Kriterien die Modellversuchsschulen und die entsprechenden Klassen und Probanden ausgewählt wurden. Die fehlenden Beschreibungen der WBL in Bezug auf die Quellenauswahl wirken sich auf die Aussagefähigkeit der Informationen insgesamt aus, da Außenstehende nicht beurteilen können, ob die Informationen, die die Evaluation hervorbringt, durch einseitige Quellenauswahl beeinflusst wurden und daher evtl. nicht ausreichend sind, um die Evaluationsfragestellungen zu beantworten.

G5 – Valide Informationen: Die Ausführungen zur Evaluation im zweiten Zwischenbericht beziehen sich ausschließlich auf die formative Evaluation und sind demnach nicht geeignet, den Einsatz der Modellversuchsvarianten zur Förderung von Schlüsselqualifikationen zu beurteilen (ZB 1993: 52ff).

Im Abschlussbericht wird lediglich für das Instrument 'Beziehungsfenster', das das erreichte Maß an 'Denken in Zusammenhängen' ermittelt, dokumentiert, welcher Anstieg im Bereich dieser Qualifikation bei den Schülern einer Klasse verzeichnet werden konnte und warum die WBL dieses Evaluationsinstrument für valide hält: Es handele sich um ein in der pädagogischen Forschung bekanntes Verfahren, bei dem Differenziertheit und Strukturiertheit der mentalen Vorstellungen der Auszubildenden bezüglich betrieblicher Zusammenhänge festgestellt werden können (vgl. AB 1994: 28). Informationen, inwiefern die WBL das Instrument aktualisiert und diese aktualisierte Fassung einer Validitätsprüfung unterzogen hat, sind den Berichten nicht zu entnehmen.

Für die anderen Modellversuchsvarianten werden keine Aussagen in den Berichten festgehalten, warum die Evaluationsverfahren ein- und wie diese umgesetzt und

welche Ergebnisse erreicht wurden; Angaben zur Validität fehlen außerdem. Durch die fehlende Operationalisierung der Evaluationsziele können zudem die Schlussfolgerungen nur den Charakter von 'Eindrücken' erhalten: "Aus der Identifikation mit dem Ladenprojekt resultiert auch ein Zuwachs an Eigeninitiative und Selbstverantwortungsbereitschaft" (AB 1994: 27).

G6 – Reliable Informationen: Die Berichte enthalten nur begrenzt Angaben über mögliche systematische Faktoren, die die Ergebnisse verändern könnten. So wird erst im Abschlussbericht festgestellt, dass es problematisch sei, dass die Anfangs- und Endevaluation, bei der die Methode der Befragung zum Einsatz gekommen ist, von unterschiedlichen Beurteilern vorgenommen wurde. Gleiches gilt für die Auswertung von Videoaufzeichnungen durch unterschiedliche Personen. Darüber hinaus gibt es in den Berichten keine Reliabilitätsinformationen, die in Bezug auf die beteiligten Schülergruppen und die Art und Weise, wie die Verfahren zur Informationsgewinnung in der modellversuchsspezifischen Evaluation benutzt werden, aussagefähig sind. Hinweise auf Fluktuationen in den Modellversuchsgruppen resp. Kontrollgruppen sind den Berichten außerdem nicht zu entnehmen.

G8 – Analyse quantitativer Informationen: Die Auswertung der Befragungen erfolgt durch den Fachbereich Psychologie der Universität Dortmund. Die Mitarbeiter stellen die Auswertungsergebnisse mit Hilfe von Tabellen dar. Diese Tabellen lassen die Ergebnisse des Vor- und Nachtests, unterteilt in Experimentalgruppe und Kontrollgruppe, sowie die Steigerungsraten erkennen. Eine Aufbereitung mittels Grafiken fehlt. Die angekündigte Auswertung der Falldarstellung mittels selbst entwickeltem Punkteschema wird in den Berichten nicht abgedruckt. Eine Interpretation der Ergebnisse unterbleibt nahezu vollständig, so dass die eingesetzten Evaluationsmethoden, die die quantitativen Daten ermittelten, nur begrenzt rechtfertigen können.

G9 – Analyse qualitativer Daten: Die gesammelten qualitativen Daten stammen aus Beobachtungen und Reflexionen und beziehen sich auf die einzelnen zu fördernden Schlüsselqualifikationen und können in diesem Sinne als kategorisiert angesehen werden. Die Ergebnisse werden innerhalb der Bezüge der durchgeführten Projekte dargestellt. Ob sie geeignet sind, die Fragestellungen zu beantworten, kann schon wegen fehlender Zieloperationalisierung nur gemutmaßt werden. Eine Analyse der Beobachtungen findet außerdem nicht statt, vielmehr werden diese lediglich dokumentiert, so dass die thesenartigen Empfehlungen sich auf eine unstrukturierte Datenmenge beziehen. (vgl. AB 1994: 69ff). Diese Ergebnisse sind nicht geeignet, Auskunft darüber zu geben, ob Schlüsselqualifikationen durch die initiierten Lernorganisationen gefördert werden können.

G10 – Begründete Schlussfolgerungen: Schlussfolgerungen der WBL in Bezug auf die Förderung von Schlüsselqualifikation, die mit Hilfe der entwickelten lernorganisatorischen Bedingungen weiterentwickelt werden konnten, werden nicht entwicklungslogisch, d. h. in Zusammenhang mit den eingesetzten Verfahren und in Bezug auf das Bedingungsfeld und die Modellversuchsvarianten, dargestellt. D.

h. die Schlussfolgerungen werden nicht auf Situationen, Zeiträume, Personen, Zusammenhänge und Zwecke bezogen, die den Modellversuchsvarianten entsprechen und für die Evaluationsergebnisse anwendbar sind. Der Abschlussbericht enthält lediglich Thesen, die nicht auf die eingesetzten Evaluationsmethoden zurück zu führen sind (vgl. AB 1994: 32).

G12 – Metaevaluation: Eine Metaevaluation ist nach Lage der Berichte nicht durchgeführt worden.

Resümee Standardbereich G

Das Niveau der Ausprägungsgrade der G-Standards macht deutlich, dass die verwendeten Methoden der Evaluation zwar geeignet sind, den überwiegenden Teil der Ausgangsthese zu belegen (Die Förderung und Entwicklung von Schlüsselqualifikationen ist abhängig von handlungsorientierten Lernprozessen), allerdings fehlen nachvollziehbare Analysen der gesammelten Daten, die in entsprechende Schlussfolgerungen münden, so dass die Güte und/oder Verwendbarkeit des evaluierten Modellversuches nicht vollständig eingeschätzt werden kann. Generell ist es das Fehlen verschiedener Informationen bzw. Arbeitsschritte, das zu der o. a. Einschätzung führt:

Beim Standard G1 (MV-Dokumentation) wird nicht deutlich, welche Änderungen im Laufe des Modellversuchs vorgenommen werden mussten. Im Hinblick auf den Standard G2 (Kontextanalyse) ist festzustellen, dass arbeitsmarkt- und betriebswirtschaftliche Aspekte hervorgehoben werden, allerdings werden weitere einflussnehmende Faktoren, wie parallel laufende Aktivitäten oder die bildungspolitische Situation der Hansestadt im Vergleich zu den anderen Bundesländern nicht aufgezeigt. Die Einstufung des Standards G3 (Beschreibung von Zielen und Vorgehen) in den Ausprägungsgrad '2' wird begründet mit der Formulierung des Evaluationszieles (ZB 1992: 23f, AB 1994: 26f, 69f) und dem nachvollziehbaren Vorgehen bei der Evaluation; kritisch ist jedoch die fehlende Operationalisierung des Formalzieles (Eingangshypothese) hervorzuheben.

Die Einschätzung der Validität der Schlussfolgerungen und der Reliabilität der eingesetzten Verfahren (G5 und G6) muss unter dem arithmetischen Mittel verbleiben, da auch hier aus den Dokumentationen die entsprechenden Informationen nicht oder nur zu einem geringen Teil entnommen werden können. Eine Analyse der gesammelten Information (G8, G9) unterbleibt bzw. ist nur zu einem Teil in die Berichte aufgenommen, so dass auch Schlussfolgerungen (G10) nur thesenhaft dargestellt werden.

Die Anwendung des Kongruenzschemas zur Feststellung der gegenseitigen Abhängigkeit von Modellversuchsaufbau inkl. Evaluationsverfahren und Verfahren der Zielermittlung findet nach dem ersten Prüfaspekt (Nennen von Evaluationszielen) ein Ende, da diese nicht operationalisiert werden.

Zwischenergebnis – Analyse und Interpretation der Ergebnisse:

Die Ausprägungen der Nützlichkeits- und Korrektheitsstandards liegen im mittleren Bereich und tragen somit zu einer Qualitätssicherung des Modellversuches bei.

Positiv ist die (wenn auch schwierige) Nachvollziehbarkeit des Evaluationsvorgehens hervorzuheben. Diese Art der Transparenz findet sich in den Ausprägungen der G-Standards wieder und gewährleistet somit die ansatzweise Transferierbarkeit der Ergebnisse.

Die WBL verweist ihm Rahmen ihrer Dokumentationen auf verschiedene Interessenkonflikte. Diese werden jedoch nicht auf ihre Ursachen zurückgeführt, so dass Änderungen hätten vorgenommen werden können. Vermutlich sind diese Schwierigkeiten auf ein zu geringes Maß an Transparenz im MV-Verlauf (siehe N1) zurückzuführen.

Kritisch ist die Undurchsichtigkeit der Dokumentationen zu werten. Es ist äußerst schwierig, den Ausführungen zu folgen und wichtige Informationen zu entnehmen.

Darüber hinaus fehlt eine Operationalisierung der Evaluationsziele, so dass eine finale und nachvollziehbare Ergebnisprüfung offen bleiben muss. Hierdurch wird die Qualität des MVs und seine Leistungen für das Praxisfeld in Frage gestellt.

6.5 Darstellung und Evaluation des Modellversuchs SOL

6.5.1 Beschreibung des Modellversuchs

6.5.1.1 Allgemeine und strukturelle Angaben

Der Modellversuch 'Selbstorganisierte Lernprozesse und neue Lernwelten in der beruflichen Bildung' ist in der Zeit vom 01.09.1997 bis 28.02.2000 im Bundesland Hessen durchgeführt worden. Die Leitung des Modellversuchs oblag dem Hessischen Landesinstitut für Pädagogik (HeLP).

Die Modellversuchsaktivitäten, an denen über 20 Schulpraktiker beteiligt waren, erstreckten sich auf etwa 90 Projekte in unterschiedlichen Berufsfeldern und berufsbildenden Schulen. Beteiligt waren zwei kaufmännische Schulen und zwei Schulen der gewerblichen Fachrichtung sowie ein Studienseminar für berufliche Schulen (vgl. ZB 2 1999: 16ff u. AB 2000: 51).

Die wissenschaftliche Begleitung wurde vom Institut für Wirtschaftspädagogik der sozial- und wirtschaftswissenschaftlichen Fakultät der Universität Bamberg unter Leitung eines Universitätsprofessors wahrgenommen. Zur Abstimmung zwischen den Modellversuchsbeteiligten und –projekten waren drei Projektkoordinatoren tätig.

6.5.1.2 Modellversuchsziele und –inhalte/Tätigkeiten der WBL auf Modellversuchsebene

Ausgehend von zwei unterschiedlich verorteten Entwicklungstendenzen, gemeint ist einerseits der technologische Wandel und die sinkende Halbwertszeit des Wissens, die geänderte Anforderungen an die Mitarbeiter stellen, andererseits die Änderungen im Bildungssystem, die die Einzelschule zum Mittelpunkt von Reformbemühungen machen, wurde ein neues Bildungs- und Lernverständnis sowie die innovative Organisation schulischer Prozesse gefordert (vgl. ZB 1 1998: 10f). In diesen Zusammenhang ist die Initiierung des Modellversuchs einzuordnen:

Formalziel des Modellversuchs waren Ergebnisse, die 'Leitvorstellungen und Empfehlungen' darstellen und neue Gestaltungsräume durch die Informations- und Kommunikationstechniken im Hinblick auf selbstgesteuertes Lernen eröffnen sollten. Im Rahmen der Entwicklung eines Konzepts zur Schulentwicklung war es dabei Gegenstand des Modellversuchs, Lernprozesse zu initiieren und zu gestalten (vgl. ZB 1 1998: 11). Hierbei sollten Simulationsmodelle für unterschiedliche Berufsfelder entwickelt werden, die vor allem den 'Einsatz von neuen Informations- und Kommunikationstechniken im Unterricht' berücksichtigen (vgl. ZB 2 1999: 14).

Darüber hinaus sind sechs erkenntnisleitende Fragen entwickelt worden, die sich mit der Profilbildung an Berufsschulen im Bereich IuK, der Veränderung der Lernkultur, der Kooperation von Modellversuch und zweiter Phase der Lehrerbildung sowie der Gestaltung von transferfähigen Prozessen und Ergebnissen aus dem Modellversuch beschäftigten. Diese Leitgedanken sind mit Unterstützung von Mitarbeitern der wissenschaftlichen Begleitung in 'standortübergreifenden Arbeitsgruppen integrativ' erarbeitet worden (vgl. AB 2000: Vorbemerkungen).

Die wissenschaftliche Begleitung selbst widmete sich drei Aufgabenbereichen, die sich auf die Implementation selbstorganisierten Lernens in Unterrichtsprojekten (Beratungs- und Gestaltungsfunktion auf der Objektebene; Anm. Verf.), auf die Datenerhebung und –auswertung ausgewählter Fragen des Modellversuchs (Evaluationsfunktion s. u.) und auf die Moderation von Diskussionsprozessen zu zentralen Fragen des Modellversuchs (Beratungs- und Gestaltungsfunktion auf der Objektebene; Anm. Verf.) bezogen.

Im Rahmen der Implementationsaufgabe der WBL hatte diese Kontakt zu den beteiligten Schulpraktikern aufgenommen, um Unterrichtsentwürfe zu diskutieren und Unterrichtsprojekte vor Ort zu besuchen. Darüber hinaus hatte die WBL ihre Beratungsfunktion durch die Moderation von Diskussionsprozessen unterschiedlicher Themengebiete in durchgeführten Workshops wahrgenommen.

Für den Aufgabenbereich 'Datenerhebung und –auswertung' wurden von der WBL Fragebögen für die unterschiedlichen Zielgruppen (Schüler und Lehrer) entwickelt, um Erkenntnisse in Bezug auf das Ausmaß der Umsetzung selbstorganisierten Lernens bzw. im Hinblick auf Schwierigkeiten bei der didaktischen Umset-

zung des SoL-Ansatzes zu erhalten (summative Evaluation; siehe Kap. 6.4.2; vgl. AB 2000: 51).

Die Befragungen sind zum Teil prozessbegleitend (bei kürzeren Projekten) bzw. am Anfang und am Ende (bei Projekten bis zu einer Laufzeit von sechs Monaten) des jeweiligen Unterrichtsprojektes durchgeführt worden. Inwiefern die Ergebnisse hieraus in den Modellversuchsprozess zurück gemeldet wurden und zu einer Veränderung des Modellversuchsverlaufes geführt haben (formative Evaluation), ist den Dokumentationen nicht zu entnehmen.

6.5.1.3 Fortbildungen/Tagungen/modellversuchsbegleitende Transferaktivitäten

Im Zeitraum September 1997 bis April 1999 haben dreizehn Tagungen zu folgenden Themen stattgefunden: Aufnahme der MV-Arbeit, Schwerpunktsetzungen im MV, Entwicklung eines Logos, Vereinbarungen mit der WBL, Schwerpunktbildung der Arbeit in den Schulen, Analyse von Elementen selbstorganisierten Lernens, Reflexion stattgefundenener Tagungen, Unterrichtshospitation, Lernfeldkonzept und SoL, Fortbildungskonzept (ZB 1998: 94; ZB 1999: 146).

Darüber hinaus ist eine Klausurtagung unter Leitung des Teams der WBL, unterteilt in drei Workshops (Profilbildung/Lernkultur an den Schulen, Bedeutung von IuK-Techniken/Neue Medien für selbstorganisierte Lernprozesse, Transfer von Erkenntnissen aus dem Modellversuch/Lehrer(innen)ausbildung) durchgeführt worden (vgl. AB 2000: 69).

Eine angekündigte, über die in den Berichten dargestellten Ergebnisse hinaus gehende Detailanalyse, die in einer wissenschaftlichen Publikation veröffentlicht werden soll, ist noch nicht erschienen (vgl. AB 2000: 54).

6.5.1.4 Ergebnisse des Modellversuchs aus Sicht der WBL

Die Ergebnisse des Modellversuchs werden entsprechend der auf dieser Ebene formulierten Zielsetzungen dargestellt, dazu werden die Modellversuchsziele zwecks Operationalisierung durch 'Thesen' (in Frageform) untersetzt, auf die im Abschlussbericht ausführlich Stellung bezogen wird. Dem Abschlussbericht sind Ergebnisse zu folgenden Leit-Fragestellungen zu entnehmen:

- Wie lässt sich an Berufsschulen im Bereich der Informations- und Kommunikationstechniken eine Profilsetzung realisieren?
- Welche Veränderungen hinsichtlich der Lernkultur können die Informations- und Kommunikationstechniken an Berufsschulen bewirken?
- Welche neuen Gestaltungsräume eröffnen multimediale und telekommunikative Systeme für selbstorganisierte Lernprozesse?
- Welchen Beitrag können die Informations- und Kommunikationstechniken für die Planung und Realisierung von simulativen Lernprozessen leisten?

- Wie können der Modellversuch und die zweite Phase der Lehrerbildung miteinander kooperieren, damit es zu einem Austausch und zu Synergieeffekten kommt?
- Wie müssen die Prozesse und Ergebnisse aus dem Modellversuch gestaltet bzw. aufbereitet sein, damit der Austausch mit den Modellversuchsschulen, dem Studienseminar und der Transfer auf Landes- und Bundesebene möglich wird?

Durch die Stellungnahme zu den in Frageform formulierten 'Thesen' werden Lösungsvorschläge zu den Problem-(Ziel-)bereichen unterbreitet, die als Erkenntnisse aus der Modellversuchsarbeit angesehen werden können (siehe AB 2000: 15 bis 50). Ein wertendes Resümee, das Auskunft über das Maß der Zielerreichung gibt, wird nicht dargestellt.

6.5.2 Darstellung des summativen Evaluationsansatzes der WBL im Modellversuch SoL

Die Datenerhebung und –auswertung, die die WBL als einen Bestandteil ihrer Konzeption der Begleitung des Modellversuchs beschrieben hat, wurde durch drei Erhebungswellen durchgeführt. Die Zielsetzung der Datenerhebung und -auswertung ist als Evaluation des modellversuchsbezogenen SoL-Leitkonzepts beschrieben worden und kann daher gleichsam als evaluationsbezogene Zielsetzung angesehen werden:

Die angesprochenen Erhebungen fanden im Sommerhalbjahr 1998 und im Frühjahr/Sommer 1999 sowie gegen Schuljahresende 1998/1999 statt. Zielgruppe der Befragungen waren am Modellversuch beteiligte Schüler (ergänzt um Kontrollgruppe) und Lehrer.

Bezugspunkte für die Konstruktion der Schüler-Fragebögen waren, in Anlehnung an Sembill, neun Merkmalsbereiche selbstorganisierten Lernens: 'Lernen als geplantes Handeln' stand hierbei im Mittelpunkt. Um diesen Mittelpunkt breitete sich ein vierdimensionales Feld aus, das durch (neun) Merkmalsbereiche dimensioniert wurde (vgl. ZB 1998: 21). Ziel war die Beantwortung erkenntnisleitender Fragen (Zielsetzung auf Ebene der Evaluation), die die WBL mit Hilfe zweier Erhebungswellen zu beantworten suchte, in denen je zwei unterschiedliche Schüler-Fragebögen zum Einsatz gekommen sind.

Der erste Schüler-Fragebogen (SOGU/S-Selbstorganisationsgrad von Unterricht/Schülerfragebogen, erste Erhebungswelle) zielte auf den Selbstorganisationsgrad von Unterricht ab. Es sollte erfasst werden, inwiefern die Schüler wahrgenommen haben, dass zentrale Aspekte selbstorganisierten Lernens im Unterricht umgesetzt bzw. aufgetreten sind (vgl. ZB 1999: 130). Der zweite Fragebogen (zweite Erhebungswelle) 'SUKESOL – Subjektive Kompetenzeinschätzung Selbstorganisierten Lernens' sollte ermitteln, "[...] inwiefern die Schüler sich Fähigkeiten und Fertigkeiten zuschrieben, die für Selbstorganisiertes Lernen wichtig sind" (ZB

1999: 132). Darüber hinaus sollte analysiert werden, welche Umstände Lernerfolge verhindert haben.

Zeitgleich mit der zweiten (schülerbezogenen) Erhebungswelle wurden Verfahren eingesetzt, die den Wissens- und Kompetenzzuwachs, die Lernmotivationsentwicklung, die Leistungsmotivation und die Selbstwirksamkeitserwartung (bei Schülern) erfassen sollten (vgl. ZB 1999: 141). Die Datenerhebungsmethoden bezogen sich auf 'Lernerfolgsvariablen', die die komplexe Problemlösefähigkeit als zentrale Zielgröße selbstorganisierten Lernens ermitteln sollten. Daneben standen die 'vermittelnden Variablen', die die Wahrnehmung relevanter Aspekte der Lernumgebung und die subjektive Einschätzung relevanter Fähigkeiten und Fertigkeiten erfassten sowie die Eingangsvoraussetzungen im Mittelpunkt des Interesses. Es heißt, hierzu seien spezielle Verfahren entwickelt worden (vgl. ZB 2 1999: 142); diese werden in den Berichten jedoch nicht dargestellt.

Zusätzlich ist eine dritte, rein lehrerbezogene Erhebungswelle mit einem nicht näher beschriebenen Fragebogen durchgeführt worden. Im Hinblick auf die Lehrerbefragung sollte ermittelt werden, an welchen Stellen die didaktische Umsetzung des SOL-Konzepts besondere Schwierigkeiten bereitete und welche Medien in den Unterrichtsprojekten zum Einsatz gekommen sind (vgl. ZB 1999: 141).

6.5.3 Ergebnis der Untersuchung - Ausprägungsgrade der Evaluationsstandards

S	N1	N3	N4	N5	N	D1	K1	K5
A	1	1	3	1	1,5	1	1	0

S	K6	K7	K	G1	G2	G3	G4	G5
A	3	0	1,0	0	1	2	1	1

S	G6	G8	G9	G10	G12	G
A	1	1	-	0	0	0,78

S=Standard; A=Ausprägungsgrad

Tab. 8: Ergebnis der metaevaluativen Untersuchung MV SOL/Hessen

N1 – Ermittlung der Beteiligten und Betroffenen: Die in den vier Modellversuchsschulen sowie im Studienseminar beteiligten Schulpraktiker wurden auf unterschiedliche Art und Weise an den Modellversuch herangeführt, informiert und ausgewählt. Das Maß der Freiwilligkeit divergiert dabei – während Praktiker einiger Schulen per Gesamtkonferenzbeschluss über die Teilnahme entscheiden konnten, wurden an anderen Schulen die Projektteams auf Vorschlag der Schulleitung gebildet. Die in den Modellversuch einbezogenen Klassen werden hinsichtlich ihres Bildungsganges und Geschlechts beschrieben. In Einzelfällen erfolgt eine genauere Charakterisierung des Bedingungsfeldes (Eingangsvoraussetzungen, Alter; vgl. ZB 1998: 41ff).

Eine weitergehende Identifizierung sog. 'Stakeholder' unterbleibt. Im Rahmen der Beschreibung der Ausgangslage wird zwar Bezug genommen, auf die allgemeinen ökonomischen und politischen Rahmenbedingungen, in die der Modellversuch ein-

gebettet ist, eine Charakterisierung der 'Abnehmer' der in den Modellversuch einbezogenen Schüler unterbleibt jedoch. Hier hätte eine Einbeziehung von Unternehmensverbänden, Kammern usw. zu einer zielgerichteteren Evaluation führen können. Eine weitergehende Identifizierung und darauf aufbauende Gruppierung der B&B unterbleibt somit.

N3 – Umfang und Auswahl der Informationen: Die Erkenntnisse, die aus der ersten Befragung der Schüler gezogen wurden, werden im Abschlussbericht den entsprechenden Evaluations-Leitfragen zugeordnet, so dass eine zielgerichtete Auswahl der Informationen vorgenommen wurde. Allerdings ist die Darstellung wenig umfangreich und enthält auch keine Detailinformationen in Form von 'reinen' Ergebnissen aus der Fragebogenauswertung; dargestellt werden lediglich die Interpretationen der Ergebnisse. Wegen der fehlenden Auswertungsergebnisse können die Interpretationen nicht nachvollzogen werden (vgl. AB 2000: 52f). Im Hinblick auf die zweite Erhebungswelle erfolgt eine Selektion aus der Vielzahl der gewonnenen Informationen. Es wird nur die Auswertung des Lernmotivationsfragebogens dargestellt und interpretiert. Da die Motivation für selbstorganisierte Lernprozesse von grundlegender Bedeutung ist, kann hier von einer Informations-Selektion im Sinne der B&B ausgegangen werden.

Wegen der nur begrenzten Darstellung relevanter Informationen und der nur teilweisen Nachvollziehbarkeit der Interpretationen muss die Einstufung in einen höheren Ausprägungsgrad unterbleiben.

N4 – Feststellung von Werten: Die dem Modellversuch zu Grunde liegenden Wertmaßstäbe lassen sich einerseits aus dem verfolgten Konzept des selbstorganisierten Lernens implizit entnehmen, andererseits aus der die Untersuchung leitenden Eingangshypothese erkennen (vgl. AB 2000: 12). In Bezug auf die Tätigkeiten auf der Objektebene waren erkenntnisleitende Fragen Ausgangspunkt, die durch weitere, im Ansatz operationalisierte Fragen untersetzt und abschließend durch Thesen ausformuliert wurden. Die Verifizierung dieser Thesen gliedert den ersten Teil des Abschlussberichts. Darüber hinaus legt die WBL die theoretischen Grundannahmen der Erstellung der Fragebögen offen und schafft so Transparenz hinsichtlich ihrer Werteorientierung auf summativer, evaluationsbezogener Ebene. Es wird hervorgehoben, dass durch die Auswertung der Fragebögen Erkenntnisse gewonnen werden sollen, die der Entwicklung von 'Leitvorstellungen und Empfehlungen' im Hinblick auf Gestaltungsräume selbstorganisierten Lernens, beeinflusst durch Informations- und Kommunikationstechnologien, dienen und im Rahmen eines Konzepts der Schulentwicklung genutzt werden sollen (vgl. ZB 1998: 11).

N5 – Klarheit des Berichts: Die Ziele der Evaluation wurden in Form von Fragen dokumentiert (ZB 1999: 130), die mit Hilfe von drei Fragebogenarten (Schülerfragebogen, zwei Erhebungswellen plus Lehrerbefragung) beantwortet werden sollten. Eine strukturierte Auswertung, die sich zu den Evaluationszielen in Verbindung setzt, unterbleibt für die erste schülerbezogene Erhebungswelle ebenso wie für die angekündigte Befragung der Lehrer, so dass die Nachvollziehbarkeit der Interpretationen nicht ermöglicht wird. Hinsichtlich der Ergebnisse der zweiten Erhebungs-

welle werden ausgewählte Ergebnisse dokumentiert und interpretiert. Eine vollständige Darstellung unterbleibt auch hier.

Resümee Standardbereich N:

Da die Beteiligten und Betroffenen hinsichtlich ihrer Informationsbedürfnisse nicht ausreichend ermittelt werden (Standard N1), können deren Interessen nicht planmäßig berücksichtigt sein. Allerdings schafft die WBL Transparenz ihrer Grundannahmen durch Offenlegung des 'Theoriekonstrukts' - Merkmalsbereiche Selbstorganisierten Lernens (Standard N4 – Festlegung von Werten; ZB 1998: 22), so dass dieser Standard in den höchsten Ausprägungsgrad eingeordnet werden kann. Auch die Ziele der Evaluation und die Ergebnisse der empirischen Datenerhebung werden zumindest teilweise genannt (N5 – Klarheit des Berichts; ZB 1999: 132). Konsequenzen und Empfehlungen werden jedoch nur in Form von Thesen formuliert, nachvollziehbare Handlungsempfehlungen unterbleiben, so dass die Nützlichkeitsstandards insgesamt nur ein Standard-Niveau erreichen, das sich im unteren Drittel der Ausprägungsgrade bewegt.

D1 – Praktische Verfahren: Die WBL hat ihre Interessen auf ausgewählte und hinsichtlich des Erhebungszeitraumes getrennte Aufgaben fokussiert und hierzu u. a. zwei (schülerbezogene) Erhebungswellen mit Hilfe von Befragungen durchgeführt. Hierzu wurden für den Untersuchungsgegenstand 'bewährte' Fragebögen ausgewählt (vgl. AB 2000: 57). In Abhängigkeit von der Dauer der im MV durchgeführten Projekte gab es entweder eine Abschlusserhebung, eine Eingangs- und Ausgangserhebung oder eine Eingangs-, Zwischen- und Ausgangserhebung (vgl. AB 2000: 57), so dass Entwicklungszuwächse im Hinblick auf die Qualifikationen selbstorganisierten Lernens ermittelt werden könnten. Durch den Vergleich mit Kontrollgruppen ist zudem ein Indiz für den Zusammenhang der Kompetenzentwicklung mit den eingesetzten Methoden selbstorganisierten Lernens gegeben. Da die Fragebögen nicht in die Berichte aufgenommen wurden, sind fundierte Aussagen über deren Eignung nicht möglich. Es stellt sich jedoch die Frage, ob Kompetenzzuwächse durch das Ermitteln subjektiver Eindrücke festgestellt werden können. Inwiefern die 'speziellen' Verfahren zur Ermittlung der Ausprägung 'komplexer Problemlösefähigkeit' und der 'vermittelnden Variablen' zweck- und zielangemessen sind, kann wegen fehlender Dokumentationen hierzu nicht beurteilt werden. Angaben über Zeitaufwand der Datenerhebung (Zeitverbrauch für das Ausfüllen der Fragebögen) fehlen. Nach Lage der Berichte wurden die neun Merkmalsbereiche selbstorganisierten Lernens durch 25 bis 42 Kriterien untersetzt (Die Fragebögen selbst wurden nicht in die Berichte eingestellt), so dass von einem vertretbaren, wenn auch 'erhöhten' Zeitaufwand ausgegangen werden kann, der zu vertretbaren Störungen auf der Objektebene führt.

K1 – Unterstützung der Dienstleistungsorientierung: Es ist Aufgabe einer Evaluation, beabsichtigte und unbeabsichtigte Wirkungen des Programms auf die Lernenden zu erkennen (vgl. Sanders 1999: 109) und dabei über die Interessen von Pädagogen und Organisationen hinaus zu blicken. Durch das Hervorheben der Bedeutung der SOL-Kompetenzen im Hinblick auf die Anforderungen zukünftiger

Mitarbeiter in allen Dienstleistungs- und Produktionsbereichen kann diese Anforderung zumindest insofern als erfüllt angesehen werden, als dass eine Qualifikationsentwicklung zum Gegenstand des Modellversuchs gemacht wurde, die sich nicht allein an den momentanen Arbeitsmarkt- und Arbeitsplatzanforderungen orientiert. Problematisch ist jedoch einerseits das Fehlen der eingesetzten Erhebungsmittel (Fragebögen) zu werten, da hierdurch die Eignung dieser Instrumente nicht vollständig eingeschätzt werden kann, andererseits werden auch die Ergebnisse der Erhebungen nicht vollständig abgedruckt und somit dieser Teil der Dienstleistungsfunktion der Evaluationen nicht erfüllt. Das Resultat sind Evaluationsergebnisse, die nicht ausreichen, um handlungsinitiiierend für Planungsmodellversuche dienen zu können.

K5 - Vollständige und faire Einschätzung: Eine bewertete Auflistung von Stärken und Schwächen des Modellversuchs unterbleibt. Es werden lediglich Teile der beabsichtigten Erkenntnisse aus den verschiedenen Modellversuchsprojekten dokumentiert.

K6 – Offenlegung der Ergebnisse: Die Zwischen- und Abschlussberichte sind beim MK Hessen oder nach Anforderung auch beim Hessischen Landesinstitut für Pädagogik erhältlich. Der Leiter des Modellversuchs wird genannt und fungiert auch als Ansprechpartner für weitere Informationen. Darüber hinaus wird im AB auch eine URL-Kontaktadresse genannt, so dass die Berichte auch aus dem Internet bezogen werden können (vgl. AB 2000: Vorbemerkungen). Eine CD-ROM mit entwickelten Materialien rundet das Informationsangebot ab.

K7 – Deklaration von Interessenkonflikten: Es werden keine Ursachen für mögliche Interessenkonflikte der beteiligten Parteien genannt. Schriftliche Vereinbarungen, wie im Konfliktfall verfahren werden kann, sind den Dokumentationen nicht zu entnehmen.

Resümee Standardbereich K:

Der Standardbereich der Korrektheitsstandards erreicht hinsichtlich des Durchschnitts der Ausprägungsgrade eine untere Niveaustufe: Dem untersten Ausprägungsgrad sind die Evaluationsstandards K5 (Vollständige und faire Einschätzung) und K7 (Deklaration von Interessenkonflikten) zuzuordnen, da die Berichte keinen Aufschluss darüber geben, welche Stärken und Schwächen die Evaluation des Modellversuchs verursachte bzw. wie im Konfliktfall zwischen Evaluatoren und Auftraggebern verfahren werden soll. Entscheidend wäre es gewesen, bei reibungslosem Verlauf diejenigen Aspekte hervorzuheben, die diesen begünstigt haben bzw. im gegenteiligen Fall die hemmenden Faktoren zu benennen. Der Standard K6 (Offenlegung der Ergebnisse; AB 2000) erreicht den höchsten Ausprägungsgrad, da einerseits erfolgte Veröffentlichungen und Berichte angegeben und die jeweiligen Bezugsquellen genannt, andererseits, die Ergebnisse der Befragungen zumindest teilweise in den Zwischen- bzw. dem Abschlussbericht offen gelegt werden. Dem Standard K1 (Unterstützung der Dienstleistungsorientierung) ist ansatzweise nach-

gekommen, doch fehlende Informationen über Evaluationsinstrumente und lückenhafte Ergebnisdokumentationen unterstützen die Dienstleistungsfunktion nicht.

G1 – MV-Dokumentationen: Ein Arbeitsplan, der den Modellversuchszeitraum zeitlich und inhaltlich gliedert, ist den Dokumentationen nicht zu entnehmen. Ebenso fehlen Beschreibungen über Änderungen im Verlauf des Modellversuchs, so dass die Zwischenberichte und der Abschlussbericht den Charakter von Ergebnisdokumentationen einnehmen. Inwieweit die WBL ihre Tätigkeiten auch tatsächlich nach den Aufgaben einer formativen Evaluation ausgerichtet haben und auf diesem Weg eine Rückkoppelung der gewonnenen Erkenntnisse in den laufenden Modellversuchsprozess erfolgt ist, kann außerdem nicht nachvollzogen werden (vgl. Kap. 6.4.1.2).

G2 – Kontextanalyse: Die allgemeinen und strukturellen Angaben in den Berichten zum Modellversuch geben Aufschluss über charakteristische Elemente der Modellversuchsschulen. Hierin sind auch Angaben hinsichtlich der Schulstandorte, der angebotenen Bildungsgänge und über Schüler- und Lehrerzahlen enthalten; zum Teil erfolgen auch Hinweise auf die Anzahl der in den Modellversuch involvierten Schulpraktiker (vgl. ZB 1998: 30ff; ZB 1999: 16ff).

Des Weiteren erfolgt die Einordnung des Modellversuchsgegenstandes in die ökonomische und gesellschaftliche Entwicklung, die die Bedeutung von Informations- und Kommunikationstechnologien sowie des selbstorganisierten Lernens verdeutlichen soll (vgl. ZB 1998: 10ff; ZB 1999: 12ff; AB 2000: 8ff). Das weitere politische und soziale Umfeld wird nicht aufgezeigt. Parallel laufende Aktivitäten werden schulbezogen zumindest teilweise genannt (vgl. ZB 1998: 63).

G3 – Beschreibung von Zielen und Vorgehen: Die Ziele der Evaluation werden in Bezug auf die unterschiedlichen Erhebungswellen offen gelegt (vgl. ZB 1999: 130ff). Da es sich bei den Zielformulierungen u. a. um Fragen handelt, können diese als Ansatz der Operationalisierung des Formalzieles bezeichnet werden. Eine Untersetzung der Fragen in Ausprägungsgrade unterbleibt (vgl. ZB 1999: 131). Die Ziele der zweiten Erhebungswelle werden nur skizziert (vgl. ZB 1999: 141). Die Evaluationsaktivitäten werden dokumentiert. Die Erläuterungen zu den eingesetzten Fragebögen gewähren einen überblickartigen Einblick in die Intentionen des Evaluationsvorgehens, allerdings fehlt ein Rückbezug zum Modellversuchsprozess, so dass notwendige Änderungen auf der Objektebene nicht ersichtlich sind, die jedoch handlungsanleitend (Dienstleistungsfunktion – Standard K1) notwendig wären; Ergebnisse und Empfehlungen können so nicht vollständig nachvollzogen bzw. interpretiert werden. Die ausformulierten Fragebögen sind in den Berichten nicht enthalten.

G4 – Verlässliche Informationsquellen: Die WBL führte Befragungen in Experimental- und in Kontrollklassen durch. Diese Gruppen wurden nicht hinsichtlich Geschlecht, Alter und Schulabschluss gruppiert und beschrieben; außerdem wird nicht deutlich, nach welchen Kriterien die Probanden ausgesucht wurden, so dass die Aussagefähigkeit der Informationen dadurch abnimmt, dass Außenstehende

nicht beurteilen können, ob die Informationen, die die Evaluation hervorbringt durch einseitige Quellenwahl beeinflusst wurden. Eine überblickartige Beschreibung der Verfahren erfolgt (vgl. ZB 1999: 132f, AB 2000: 57). Die Evaluatoren begründen den Einsatz verschiedener Fragebögen, indem sie deren Einsatzgebiete erläutern; somit geben sie Hinweise und Begründungen für ihr eigenes Vorgehen (vgl. ZB 1999: 129ff). Allerdings fehlen in den Berichten diese ausformulierten Fragebögen; es sind lediglich allgemeine Beschreibungen zum Aufbau des Instruments vorhanden, so dass diese nicht abschließend, vor allem im Hinblick auf das potenzielle Maß der Beeinflussung durch die Probanden, eingeschätzt werden können. Das Problem der Subjektivität, das durch Umfragen bei Betroffenen entsteht, wird vermutlich nicht durch andere Evaluationsmethoden ausgeglichen; zumindest fehlen ergebnisorientierte Aussagen hierzu.

G5 – Valide Informationen: Die Auswertung der Fragebögen bringt quantitative Daten hervor; die Schüler konnten vermutlich (lediglich Schlussfolgerung; Anm. Verf.) ihre persönlichen Eindrücke den Ausprägungsgraden '1 bis 6' zuordnen. Eine überblickartige Beschreibung der Evaluationsmethoden erfolgt (siehe D1/G3). Durch die Darstellung der Auswertung wird ersichtlich, wie die Positionen der Schüler mit Punktwerten versehen wurden bzw. wie groß die Streuung der Einschätzungen zwischen den einzelnen Schülern war, so dass Aussagen über den Ausprägungsgrad der jeweiligen Merkmalsbereiche selbstorganisierten Lernens (nach Sembill) im Unterricht nachvollzogen werden können (vgl. ZB 1999: 132ff). Die Befragung der Lehrer wird jedoch ebenso wenig dargestellt (nach Ansicht der WBL fallen Lehrerbefragungen eher in den Bereich der 'Implementationsforschung' und sollen daher nicht dargestellt werden (vgl. ZB 1999: 143)) wie eine methoden- und medienbezogene Auswertung des Unterrichts, so dass die Ergebnisse der Schülerbefragungen nicht mit der Art des Unterrichts in Verbindung gesetzt werden können und Aussagen über das Erreichen der Evaluationsziele in diesem Bereich verbleiben müssen. Interpretationen der Ergebnisse des methodischen Ansatzes SOL können demnach nicht nachvollzogen werden. Die Evaluatoren heben im Hinblick auf die Validität ihrer Instrumente lediglich hervor, dass es sich bei den eingesetzten Fragebögen um selbstentwickelte und 'bewährte' Instrumente handele (vgl. AB 2000: 57). Eine Begründung für das Ausmaß der geltend gemachten Validität erfolgt nicht, d. h. es fehlen Hinweise auf andere Einsatzbereiche der Fragebögen und die Häufigkeit des Einsatzes.

G6 – Reliable Informationen: Angaben über die hier spezifischen situations- und probandenbezogenen systematischen Faktoren, die die Ergebnisse beeinflussen können, sind in den Berichten nicht enthalten; Maßnahmen zur Reduzierung von Störeinflüssen sind entweder nicht geplant oder aber in den Berichten nicht enthalten. Darüber hinaus fehlen Informationen über Fluktuationen in den Modellversuchs- bzw. Kontrollgruppen oder andere Reliabilitätsinformationen, die für die Untersuchung unmittelbar relevant sind. Mögliche Störeinflüsse auf Schülerebene hätten z. B. durch die angekündigte Befragung der Lehrer zumindest teilweise ausgeglichen werden können, d. h. Minimierung der Störfaktoren durch zusätzliche Untersuchungsschritte oder –methoden.

G8 – Analyse quantitativer Informationen: Die Ergebnisse der Schülerbefragungen werden statistisch ausgewertet und grafisch aufbereitet (ZB 1999: 133f, AB 2000: 51ff). Eine getrennte Darstellung der Ergebnisse nach Experimental- und Kontrollgruppe verbleibt. Darüber hinaus fehlen Schlussfolgerungen aus den Interpretationen heraus (im Sinne von Empfehlungen), welche Veränderungen im Modellversuchsverlauf vorgenommen werden müssten, damit ein Optimierungsprozess initiiert werden kann, denn nicht alle Merkmalsbereiche selbstorganisierten Lernens wurden auch als 'optimal' erreicht (höchster Ausprägungsgrad) eingeschätzt. Durch den fehlenden Zusammenhang von Datenerhebung (fehlende Fragebögen), Datenauswertung (Vergleich mit Kontrollgruppe) und Interpretation können die Ergebnisse die eingesetzten Evaluationsmethoden nur begrenzt rechtfertigen.

G9 – Analyse qualitativer Daten: Qualitative Daten werden nicht erhoben.

G 10 – Begründete Schlussfolgerungen: Die Ausführungen zu den Ergebnissen der Evaluation beziehen sich ausschließlich auf die Schülerbefragung (vgl. ZB 1999: 136ff) und nicht auf die Gesamtheit der beteiligten Personen (Lehrer), so dass nicht alle relevanten Daten in die Ergebnisse einfließen. Auch wird keine Verbindung zwischen den Wirkungen selbstorganisierten Lernens auf der Schülerebene mit den eingesetzten Unterrichtsmethoden hergestellt, so dass deren Eignung nicht beurteilt werden kann. Darüber hinaus wurden die Beteiligten und Betroffenen und deren Informationsbedürfnisse (siehe Standard N1) nur unzureichend ermittelt, so dass nicht sicher gestellt ist, dass die Ergebnisse der Evaluation die notwendigen Informationen darstellen.

G12 – Metaevaluation: Eine Metaevaluation wurde nach Lage der Berichte nicht durchgeführt.

Resümee Standardbereich G:

Der geringe durchschnittlicher Ausprägungsgrad des Standardbereiches der Genauigkeitsstandards ist auf die Einstufung von fünf Standards in den Ausprägungsgrad '1' zurückzuführen.

Der Standard G1 (MV-Dokumentation) wird wegen der fehlenden Dokumentation der Entwicklungen und deren anschließenden Implementation in den Modellversuchsaufbau ebenso wie der Evaluationsstandard G 12 (Metaevaluation), wegen der nicht vorhandenen Metaevaluation, in den schwächsten Ausprägungsgrad eingestuft. Begründete Schlussfolgerungen (Standard G 10) müssen schon wegen der nicht durchgängigen Operationalisierung der Evaluationsziele unterbleiben. Die mangelnde Ermittlung der Beteiligten und Betroffenen (Standard N1) führt darüber hinaus dazu, dass Empfehlungen nicht zielgerichtet ausgesprochen werden können; daraus resultiert die Einstufung auch dieses Standards in den Ausprägungsgrad '0'.

Mit Ausnahme des Standards G3 (Ausprägungsgrad ,2'; ZB 1999: 122ff) und G9 (kann nicht angewendet werden) erreichen alle anderen Evaluationsstandards den Ausprägungsgrad '1'.

Im Hinblick auf die Validität (Standard G5) der eingesetzten Evaluationsinstrumente reicht der Hinweis auf die 'Bewährung' der Fragebögen nicht aus. Hier hätten Informationen über die bisherigen Einsatzbereiche und die Angleichung bzw. Ausrichtung auf den hier spezifischen Anwendungsbereich zu einer Steigerung der Validität geführt. Auch die Reliabilität (Standard G6) leidet unter der mangelnden Aussagefähigkeit der Informationen: Die Ergebnisse der Lehrerbefragung werden nicht dokumentiert und die Erkenntnisse der Schülerbefragung werden nur selektiert dargestellt. Darüber hinaus sind keine Maßnahmen ergriffen worden, die generellen Nachteile der Methode der Befragung durch andere Untersuchungsmethoden auszugleichen. Als Ergebnis des niedrigen Niveaus der Genauigkeitsstandards ist festzustellen, dass die gesammelten Daten nicht ausreichen, die Güte und/oder Verwendbarkeit des evaluierten Modellversuchs vollständig einschätzen zu lassen.

Die Anwendung des Kongruenzschemas zur Feststellung der gegenseitigen Abhängigkeit von Modellversuchsaufbau inkl. Evaluationsverfahren und Verfahren der Zielermittlung kann vollständig angewendet werden:

Die Ziele der Evaluation werden in Form von Fragen dokumentiert (vgl. ZB 1999: 132), so dass diese schon als einen ersten Schritt der Operationalisierung angesehen werden können. Darüber hinaus sind die Ziele getrennt nach Erhebungswellen dokumentiert und es wird deutlich, mit welchen Erhebungsinstrumenten, gemeint sind die selbstentwickelten Fragebögen 'SOGU/S und SUKESOL, die gewünschten Daten ermittelt werden sollen. Da es sich bei den Erhebungsdaten um Kompetenzausprägungen handelt, hat die WBL ex-ante- und ex-post-Befragungen in Modellversuchsklassen (und auch in Kontrollklassen) durchgeführt, um durch Vergleich der Daten den Kompetenzzuwachs ermitteln zu können. Problematisch ist das Fehlen der Fragebögen in den Modellversuchsberichten sowie die nur überblickartigen Ergebnisdarstellungen und –interpretationen zu werten.

Die vorhandenen Informationen zu den Evaluationsinstrumenten stellen Indizien dar, die die Vermutung nur bedingt unterstützen, dass die Instrumente geeignet sind, die Kompetenzen resp. deren Entwicklung zu prüfen. Allerdings sind auch auf Grund der lückenhaften Informationslage (s. o.) keine abschließenden Einschätzungen zu den Evaluationsinstrumenten und im Hinblick auf den Erfolg des Modellversuchs aus inhaltlicher Sicht und der Angemessenheit der getroffenen Aussagen und Interpretationen möglich. Das folgende Zwischenergebnis stellt daher nur Erkenntnisse dar, die sich auf eine nicht vollständige Modellversuchsdokumentation beziehen müssen:

Zwischenergebnis – Analyse und Interpretation der Ergebnisse:

Die Auswertung der Schülerbefragungen und die dargestellten Modellversuchsergebnisse (vgl. AB 2000: 15ff) unterstützen die These, dass auf der Ebene des Mo-

dellversuchs die Verbindung von selbstorganisierten Lernprozessen mit multimedialen und telekommunikativen Techniken gelungen ist. Für die Evaluationsebene wäre eine differenziertere Auseinandersetzung sowie ausführlichere Interpretationen und Schlussfolgerungen dringend erforderlich, um fundierte Aussagen zur Zielerreichung tätigen zu können.

Die schwache Ausprägung der Nützlichkeitsstandards macht deutlich, dass sich die Evaluatoren nicht ausführlich mit ihren Adressaten vertraut gemacht, ihre Informationsbedürfnisse ermittelt und eine auf diese Bedürfnisse zugeschnittene Evaluation geplant haben, so dass die Analyse und Bewertung der Ergebnisse der Untersuchung Erkenntnisse hervorbringt, die nicht zielgruppenadäquat sein können. Wenn sie für den Auftraggeber und die anderen Betroffenen und Beteiligten von Interesse sind, dann handelt es sich dabei um eine zufällige Übereinstimmung von Informationsbedürfnissen und Ergebnissen der Evaluation. Durch die Korrelation mit dem Standard K1 zeigen sich die Wirkungen der nicht ausreichenden Ermittlung der Beteiligten und Betroffenen auch in diesem Bereich. Die Standards K7 und G1 beziehen sich vor allem auf dokumentarische Pflichten, die hier nicht erfüllt wurden. G9 konnte nicht eingeschätzt werden, da die Anwendung dieser Form der Befragung quantitative Daten zum Ergebnis hatte. Des Weiteren sind vor allem die Standards G5 (Validität) und G6 (Reliabilität) sowie G8 (Analyse quantitativer Daten), die einer generellen Verbesserung bedürfen, um zu einer Qualitätssteigerung der Evaluation zu gelangen.

Positiv hervorzuheben sind jedoch die erstellten Arbeitsmaterialien, so dass zumindest für die Modellversuchsebene Ergebnisse vorhanden sind und somit begrenzt transferfähige Produkte erzeugt wurden.

6.6 Darstellung und Evaluation des Modellversuchs EUWAS

6.6.1 Beschreibung des Modellversuchs

6.6.1.1 Allgemeine und strukturelle Angaben

Der Modellversuch 'Integration einer Europäischen Dimension in die kaufmännische Ausbildung zur/zum Euro-Wirtschaftsassistentin/ Euro-Wirtschaftsassistent' ist in der Zeit vom 01.08.1996 bis zum 31.07.1999 im Bundesland Nordrhein-Westfalen durchgeführt worden. Die regionale und bezirksübergreifende Koordination des Modellversuchs erfolgte durch die Bezirksregierung Detmold als Projektträger. Das Ministerium für Schule und Weiterbildung koordinierte Modellversuchsaktivitäten auf Landesebene.

An dem Modellversuch waren drei Schulen an den Standorten Bielefeld, Düsseldorf und Köln beteiligt. Hier wurde, im Zusammenhang einer kaufmännischen dualen Berufsausbildung, die 640 Unterrichtsstunden umfassende Zusatzqualifikation 'Euro-Wirtschaftsassistent' etabliert. Der Unterricht begann frühestens mit dem zweiten Ausbildungsjahr, erstreckte sich auf die Dauer von zwei Jahren und wurde an zwei Abenden zu je vier Unterrichtsstunden angeboten (vgl. ZB 1 1996/97: 8). Die Zielgruppe dieser Zusatzqualifikationen variierte je nach Schulstandort. Generell angesprochen waren alle diejenigen Auszubildenden in Ausbildungsgängen des

dualen Systems, die internationale Beziehungen zu anderen Unternehmen wahrnehmen könnten. Ausgeschlossen waren somit z. B. sog. 'Helferberufe'.

Dem Arbeits- und Zeitplan sind vier Phasen zu entnehmen: Die Problempräzisions- und Konstituierungsphase, in der an den Schulstandorten eine Zielpräzisierung sowie die Konsensfindung über inhaltliche Arbeitsschwerpunkte u. a. stattfand. Nachfolgend wurde eine Konzeptualisierungs- und Entwicklungsphase initiiert, die vor allem der Curriculumentwicklung diene. In der Erprobungs- und Evaluationsphase fand die Curriculumerprobung- und revision, die Erprobung von Organisationsmodellen des Unterrichts und des Praktikums sowie die Evaluierung der Konzepte statt. Die Abschlussphase diene vor allem der Dokumentation und Aufbereitung der Modellversuchsmaterialien (vgl. ZB 1996/97 u. 1997/1998 u. AB 1999: 6f).

Die wissenschaftliche Begleitung wurde vom Institut für Berufs- und Wirtschaftspädagogik der Universität Köln wahrgenommen, an der vier Personen, darunter ein Universitätsprofessor und ein promovierter wissenschaftlicher Angestellter, mitwirkten.

6.6.1.2 Modellversuchsziele- und inhalte/Tätigkeiten der WBL auf Modellversuchsebene

Generelles Ziel des Modellversuchs war die Entwicklung eines Curriculums für die Zusatzqualifikation 'Euro-Wirtschaftsassistent' zur Berufsausbildung im Dualen System.

Zentraler Ausgangspunkt für die Konzeption des Modellversuch war dabei die bildungspolitische Forderung nach der 'Integration einer europäischen Dimension' in Lehrpläne. Zurückzuführen sei dies auf die Einführung des europäischen Binnenmarktes, die Auswirkungen auch auf kleinere und mittlere Betriebe in Bezug auf die Arbeitsbedingungen und Qualifikationsanforderungen mit sich ziehe (vgl. ZB 1996/97: 58).

Gegenstand des Modellversuchs war daher die 'Entwicklung und Erprobung' eines Curriculums für die Ausbildung zur/zum Euro-Wirtschaftsassistent/in, dessen Fokus auf die Qualifikationen einer Sachbearbeiterfunktion in Unternehmen kleiner und mittlerer Größe gerichtet war. Kennzeichen dieser Unternehmen ist eine europäische Orientierung in allen Bereichen der Beschaffungs- und Absatzmärkte.

Für die Entwicklung dieses Curriculums waren zwei Fixpunkte entscheidend: Einerseits war bei der Erarbeitung dieses 'Lernfeld-Curriculums als Zusatzqualifikation zur kaufmännischen Erstausbildung' die Kombination schulischer und betrieblicher Lehr-Lernprozesse zu berücksichtigen, andererseits waren wirtschaftliche, fremdsprachliche und politisch-kulturelle Fachperspektiven als Bezugspunkte zur 'europäischen Dimension' aufzunehmen (vgl. ZB 1996/97: 58 und AB 1999: 115).

Mit diesen Leitgedanken (primäre Zielsetzung), die auch als Beitrag zur ‚Differenzierung und Stabilisierung des Dualen Systems‘ angesehen wurden, waren weitere zwei (unmittelbare) Zielsetzungen verbunden:

Ausgehend von der Überzeugung, dass ein Curriculum solange wirkungslos bleibe, bis auch Handreichungen für diejenigen entwickelt seien, die nicht am Modellversuch beteiligt werden konnten, waren Unterrichtsmaterialien und –medien zu entwickeln, die eine Unterrichtsgestaltung im didaktisch-methodischen Sinne der Handlungsorientierung unterstützten.

Drittes Ziel war die Integration eines Auslandspraktikums, das die 'europäische Dimension' in Modellen der Zusammenarbeit zwischen Bildungsträgern und Betrieben im In- und Ausland ausdrückt.

Darüber hinaus stellten Reflexionen zum Transfer der Erkenntnisse und Überlegungen zu Lehrerfortbildungsmaßnahmen mittelbare Zielsetzungen des Modellversuchs dar (vgl. ZB 1996/97: 59).

Dem Grundverständnis der Mitglieder der WBL entsprach die unmittelbare Einbindung der beteiligten Schulpraktiker, die sich z. B. in einem ‚akzeptanzsichernden internen Evaluationsprinzip‘, also u. a. in der internen Prüfung von Modellversuchs-ideen, -ansätzen und –ergebnissen, äußerte. Darüber hinaus war die Berücksichtigung des ‚individuell Machbaren und individuell Gewünschten‘ sowie die Wahrung relativer Autonomie der Schulstandorte im Sinne eines ‚bottom-up-Prinzips‘ Kennzeichen des Selbstverständnisses der WBL (Arbeitsprinzipien; vgl. ZB 1996/97: 61).

Der Untersuchungsansatz der wissenschaftlichen Begleitung wurde von dieser als 'qualitativ angelegte Modellversuchsforschung' gekennzeichnet. In diesem Zusammenhang formuliert sie drei Aufgaben:

Die dokumentarische Aufgabe bestand in der Erfassung von Prozessen zwecks Nachvollziehbarkeit von entstandenen und nicht entstandenen, potenziellen Modellversuchsergebnissen. Mit Hilfe formaler Kriterien, die in den MV-Dokumentationen nicht dargestellt werden, erfolgte die Verfahrensdokumentation als Verschriftlichung der 'teilnehmende Beobachtung' im Anschluss an die Treffen der einzelnen Modellversuchsgruppen. Hier flossen subjektive Eindrücke einzelner Teammitglieder und im Rahmen einer internen kommunikativen Validierung gemeinsam wahrgenommene und interpretierbare 'Fakten' ein.

Im Rahmen der beratenden Aufgabe wurde das 'aktive Zuhören' als methodische Variante eingesetzt, das vor allem im Zurückspiegeln von Fragen Ausdruck findet. Hierdurch ergab sich eine 'Gratwanderung', die sich in der möglichen Auflösung der Distanz zum Modellversuchsfeld ausdrückte. Die Berücksichtigung der o. g. Arbeitsprinzipien und die (begrenzt vorhandene) Möglichkeit der Diskussion von Arbeitsberichten wurde dabei als Regulativ angesehen (vgl. ZB 1996/97: 64).

"Die evaluativ-konstruktive Aufgabe lag in der Untersuchung der Wirkungen oder Nicht-Wirkungen von 'benutzten Theorien' (Theorieanwendung) oder der Fortschreibung und Erweiterung 'bestehender Theorien' (Theoriebildung)" (ZB 1996/97:

64). Diese Funktion der WBL stand in enger Verbindung mit der Beratung, etwa durch die Präzisierung von Begriffen und der kontextbezogenen Rezeption von Literatur. Darüber hinaus diente eine Schülerbefragung im Frühjahr 1998 der internen Revision des Curriculums; die Erkenntnisse hieraus sind in den Modellversuchsprozess zurück geflossen, so dass hier eine formative Evaluation durchgeführt wurde (vgl. ZB 1997/1998: 27ff).³⁸

Angaben über Tagungen/Fortbildungen oder transferbegleitende Aktivitäten sind in den Berichten nur begrenzt enthalten: Der zweite Zwischenbericht weist unter dem Gliederungspunkt 'Modellversuchsaktivitäten im Berichtsjahr' auf vier eintägige und eine zweitägige 'Besprechung' mit unterschiedlichen Beteiligten hin (vgl. ZB 1997/1998: 8).

6.6.1.3 Ergebnisse des Modellversuchs aus Sicht der WBL

Die Ergebnisse des Modellversuchs stellen einerseits das entwickelte Lernfeld-Curriculum dar, das dem Anhang des Abschlussberichts (AB 1999: 219ff) zu entnehmen ist, andererseits formuliert die WBL ihre Ergebnisse in Form von 21 Thesen. Kernpunkte mit unterschiedlichen Zielebenen hierin sind:

- Das Fachverständnis ist dem Lernfeldverständnis gewichen.
- Die Zusatzausbildung wird von allen Beteiligten uneingeschränkt befürwortet.
- Das die Lernfelder ergänzende obligatorische Praktikum wird als unabdingbar betrachtet, wobei eine Ausweitung auf mehrere Monate vorgeschlagen wird.
- Insbesondere Unternehmen mit flachen Hierarchien böten sich für den Einsatz von EUWAS-Absolventen an.
- Die Marktchancen der Absolventen werden als deutlich besser dargestellt, als die der Kaufmannsgehilfen ohne Zusatzqualifikation.
- Eine 'bessere' Position der Absolventen in den Unternehmen sei nicht gewährleistet (vgl. AB 1999: 203ff).

6.6.2 Darstellung des summativen Evaluationsansatzes der WBL im Modellversuch EUWAS

Da die Tätigkeiten im Modellversuch EUWAS als Prozess mit Einflussmöglichkeiten durch die Beteiligten organisiert sind, ergab sich nach Aussagen der WBL ein 'iterativer Ablauf', der nach einem Jahr Modellversuchsarbeit zwei Fragestellungen herausbildete, die von der WBL als 'Forschungsfragen' charakterisiert wurden:

Erstens: Welche Kriterien können angelegt, operationalisiert und gemessen werden, um den inneren Zusammenhang der zu erreichenden Kompetenzen und Inhalte in Lernfeldern zu gewährleisten? Zweitens: "Lernfelder als Ordnungsstruktur für einen Bildungsgang erfordern Möglichkeiten, einen äußeren Zusammenhang

38) Zur Darstellung der Instrumente der formativen Evaluation siehe Kap. 6.6.2. Da diese Evaluationsinstrumente auch für die summative Evaluation eingesetzt wurden, erfolgt eine gemeinsame Darstellung.

zwischen den Lernfeldern herzustellen und zu variieren. Wie können Lernfelder in ihrer Gesamtheit den Bildungsgang profilieren? Wie können Verbindungslinien zwischen Lernfeldern eine Bildungsgangdidaktik unterstützen?" (ZB 1996/97: 65)³⁹.

Um das entwickelte Curriculum zu evaluieren, wurde eine schüler und lehrerbezogene interne sowie eine externe Evaluation durch betriebliche und wissenschaftliche Experten durchgeführt. Die interne-schülerbezogene Evaluation kann auf Grund der beispielhaft in der Dokumentation dargestellten standardisierten Fragen mit dem Ziel der statistischen Auswertung (vgl. Bortz 1995: 232f, 272f) als quantitative Erhebungsmethode charakterisiert werden (vgl. ZB 1997/98: 27ff). Die Befragung der Fremdsprachenlehrer dagegen zielte auf die Erhebung subjektiver Bewertungen und selektiver Wahrnehmungen ab, was demnach als qualitative Erhebungsmethode einzustufen ist; es wurde dazu ein Interviewtyp (problemzentriertes Interview) mit einem geringen Grad an Standardisierung gewählt. Die gleiche Zuordnung trifft auf die für die externe Evaluation gewählte Methode zu. Hier wurden Gutachter (drei Führungskräfte im Personalbereich, ein Wissenschaftler) um die Beantwortung ausgewählter Fragen zum erstellten Curriculum gebeten. Die Antworten wurden einerseits aufgelistet, andererseits interpretiert dargestellt (vgl. ZB 1997/98: 47ff).

6.6.3 Ergebnis der Untersuchung - Ausprägungsgrade der Evaluationsstandards

S	N1	N3	N4	N5	N	D1	K1	K5
A	3	3	3	3	3	3	3	3

S	K6	K7	K	G1	G2	G3	G4	G5
A	3	0	2,25	3	3	2	2	1

S	G6	G8	G9	G10	G12	G
A	1	3	2	2	0	1,9

S=Standard; A=Ausprägungsgrad

Tab. 9: Ergebnis der metaevaluativen Untersuchung MV EUWAS/NRW

N1 – Ermittlung der Beteiligten und Betroffenen: Die beteiligten Personen werden auf Grund ihrer Funktion verschiedenen Gruppen zugeordnet (Projektgruppe, Sprachengruppe, Dezernenten usw.). Die Einbindung erfolgt nach dem 'bottom-up'-Prinzip, um die individuellen Wünsche zu berücksichtigen und sie an der Curriculumentwicklung aktiv teilhaben lassen zu können (vgl. ZB 1996/1997: 61f).

Entsprechend der Zugehörigkeit zum Schulstandort werden die beteiligten Modellversuchsklassen hinsichtlich Ausbildungsgang, Alter und Wohnort charakterisiert. Durch die Integration eines Auslandspraktikums sowie die generelle Ausrichtung auf kleine und mittelständige Unternehmen mit europäischer Orientierung sind auch

39) Der Begriff 'Forschungsfragen' erscheint nicht angebracht, da Tätigkeiten im Modellversuchsprozess nicht als (Berufsbildungs-) Forschung angesehen werden können (vgl. Kap. 2.1.3). In Zusammenhang mit den Erkenntnissen, die daraus gewonnen werden sollen, handelt es sich vielmehr um Fragestellungen, die geeignet sind, auf einer summativer Ebene evaluiert zu werden, um den Transfer der Ergebnisse unterstützen.

schon potenzielle Nachfrager der Absolventen identifiziert und angesprochen, so dass der Kreis der Beteiligten und Betroffenen ausgeweitet wurde. Durch die Einbeziehung der Mitglieder der Studienseminare und der Personal-Manager wird die möglichst umfassende Ermittlung der Beteiligten und Betroffenen unterstützt (siehe auch ZB 1997/1998: 171ff).

N3 – Umfang und Auswahl der Informationen: Die Berichte lassen die Entwicklung des Curriculums nachvollziehen (vgl. ZB 1996/1997: 95f, ZB 1997/1998: 67f u. 221ff, AB 1999: 203f). Außerdem beschäftigt sich die WBL schon im Entwicklungsstadium mit der Frage des Transfers und der Entwicklung von Ergebnissen als Beitrag zur Stabilisierung des Dualen Systems (vgl. ZB 1996/1997: 58f). Hinsichtlich des Transfers und der Klarheit und Wahrheit der Informationen wird der Aspekt des Umfangs explizit angesprochen und die Frage aufgeworfen, ob die Ausführungen nicht zu kurz ausfallen würden, wenn das vom BMBF geforderte Postulat 'Klarheit, Kürze und Verständlichkeit der Berichte' berücksichtigt würde. Um dennoch die Interessen von potenziellen Interessenten (Dienstleistungsfunktion – K1) berücksichtigen zu können, widmen sich die beiden Zwischenberichten den Ergebnissen der Modellversuchsebene und der Abschlussbericht ausschließlich den Erkenntnissen der WBL, was zu einer Qualitätssteigerung der Dokumentationen insgesamt beiträgt.

N4 – Feststellung von Werten: Der erste Zwischenbericht enthält standortübergreifende Arbeitsprinzipien, die die Tätigkeit der Schulpraktiker und der wissenschaftlichen Begleitung charakterisieren (vgl. ZB 1996/1997: 62). In diesem Zusammenhang stellt die WBL ihr Aufgabenverständnis heraus und erläutert ihre Funktionen im Modellversuch. Hiermit wird nicht nur ein großes Maß an Transparenz geschaffen, vielmehr macht das im ersten Zwischenbericht geschilderte Grundverständnis der Mitglieder der WBL auch ihre Werteorientierung deutlich, die ihren Ausdruck vor allem in der unmittelbaren Einbindung der Schulpraktiker und der Verbindung von 'evaluativ-konstruktiver' Tätigkeit und Beratung findet (vgl. ZB 1996/1997: 61ff; ZB 1997/1998: 27ff). Darüber hinaus kann implizit eine Werteorientierung im Sinne eines 'epistemologischen Menschenbildes' aus der Art der summativen Evaluation entnommen werden: einerseits werden hier die unmittelbar Betroffenen (Schüler und Lehrer) hinsichtlich ihrer Eindrücke befragt, andererseits werden auch die Nachfrager der zu entwickelnden Qualifikationen einbezogen, um die Akzeptanz des Lernfeldcurriculums sicher zu stellen.

N5 – Klarheit des Berichts: Der Untersuchungsansatz der wissenschaftlichen Begleitung wird im ersten Zwischenbericht skizziert und zwei 'Forschungsfragen', die als Ziele auf der (summativen) Evaluationsebene angesehen werden können, dargestellt (vgl. ZB 1996/1997: 65). Eine vertiefende Charakterisierung der Evaluationstätigkeiten erfolgt im zweiten Zwischenbericht (vgl. ZB 1997/1998: 27ff); hier wird deutlich, dass eine interne und eine externe Evaluation vorgesehen ist, auf welche Zielgruppen abgestellt wird und was diese Evaluationstätigkeiten zum Gegenstand haben (Evaluation des Lernfeldcurriculums), dass es sich um Fragebögen resp. (narrative) Interviews handelt, die als Evaluationsinstrumente zum Ein-

satz kommen sollen. Darüber hinaus geht aus den Dokumentationen hervor, wer diese Evaluationen vornimmt und welche Qualifikationen diese Personen mitbringen, so dass eingeschätzt werden kann, ob sie in der Lage sind, die Instrumente sachgerecht anzuwenden. Die Auswertung gibt umfangreich Auskunft über die Ergebnisse der Befragungen und bezieht sich dabei auf die einzelnen Frageaspekte (vgl. AB 1999: 41ff).

Resümee Standardbereich N:

Die Ausführungen der WBL sind im Hinblick auf die Informationsbedürfnisse ausführlich und nachvollziehbar. Es besteht ein hohes Maß an Transparenz. Die Evaluation der durch die WBL formulierten Ziele kann den Dokumentationen vollständig entnommen werden. Allerdings sind die Ziele lediglich als Fragen formuliert und nicht durch Ausprägungsgrade untersetzt bzw. anderweitig operationalisiert. Alle Standards können in den höchsten Ausprägungsgrad eingeordnet werden, so dass der Standardbereich insgesamt auf höchstem Niveau ausgeprägt ist. Durch die vollständige Identifikation der Beteiligten und Betroffenen, die in unterschiedlichen Funktionen in die Evaluation einbezogen werden, ist sichergestellt, dass alle Informationsbedürfnisse befriedigt werden können. Vor allem die Nachvollziehbarkeit der Evaluationstätigkeiten und die angemessene Berichtslegung trägt dazu bei, das Transferpotenzial des Modellversuchs, bezogen auf diese Standard-Ebene, zu erhöhen.

D1 – Praktische Verfahren: Die Befragung der Schüler fand in zwei Erhebungswellen mit einem Jahr Abstand statt. Auszufüllen war je ein Fragebogen, unterteilt in vier Gegenstandsbereiche, mit insgesamt 36 Fragen, die durch Ankreuzen oder Rangfolgebildung zu beantworten waren (vgl. AB 1999: 41ff). Des Weiteren hat eine Lehrerbefragung, ebenfalls mittels Fragebogen, stattgefunden: hier waren es 40 Fragen, bei denen ebenfalls durch Ankreuzen die persönliche Einschätzung deutlich gemacht werden konnte (vgl. AB 1999: 101ff). In beiden Fällen waren die Fragen durch Ausprägungsgrade untersetzt, wobei die Anzahl der Ausprägungsgrade variierte. Im Unterschied hierzu fand die Befragung der 'Fremdsprachenlehrer' mit Hilfe problemzentrierter Interviews statt, da hier die subjektive Interpretation des Lehrplans durch die entsprechenden Schulpraktiker ermittelt werden sollte (vgl. 135ff). Für die externe Evaluation durch Unternehmer und einen außenstehenden Wissenschaftler sind schwach strukturierte Interviews zum Einsatz gekommen (vgl. AB 1999: 183ff), da die Evaluatoren vermuteten, dass bei der Versendung von Fragebögen die Rücklaufquote zu gering sei, so dass die Aussagefähigkeit dieses Evaluationsteiles insgesamt nicht sicher gestellt würde.

Die Evaluationsmethoden sind im Hinblick auf die formulierten Ziele geeignet, die gewünschten Informationen zu erheben, da die Inhaltsbereiche der Befragungen darauf abzielen zu prüfen, inwiefern Lernfelder als Ordnungsstrukturen für ein Curriculum geeignet sind, den inneren und äußeren Zusammenhang des Curriculums sicher zu stellen. Unterstützt wird diese Einschätzung durch die Zielgruppenwahl der Befragten, da alle Beteiligtegruppen ausgewählt wurden. Im Hinblick auf den zeitlichen Aufwand und die daraus resultierenden Störeinflüsse auf der Objektebene sind die Evaluationsinstrumente als angemessen zu beurteilen, gerade weil das

Ausfüllen der Fragebögen resp. die Interviews keinen zu großen Zeitaufwand für die Befragten erfordern.

K1 – Unterstützung der Dienstleistungsorientierung: Die Ergebnisse der Befragungen werden ausführlich dargestellt und interpretiert, dabei werden korrespondierende Fragen gegenüber gestellt (Kreuztabelle), um Zusammenhänge aufzeigen zu können (siehe hierzu u. a. AB 1999: 48). Dies trägt dazu bei, die Wirkungen des Lernfeldcurriculums auf die Lernenden offen zu legen. Somit werden Informationen generiert, die geeignet sind, den Explikationsmodellversuch durch Modifikation in einen Planungsmodellversuch zu transferieren, ohne ersteren lediglich zu adaptieren (Dienstleistungsfunktion für die Modellversuchspraxis).

Im Hinblick auf die Bildungsziele des Ausbildungsganges bewerten die Evaluatoren auch die Abstimmung der einzelnen Lernfelder untereinander, um den Charakter der ganzheitlichen Ausrichtung des Curriculums sicher zu stellen. Hierdurch wird vermieden, dass eine Zusatzqualifikation ins Leben berufen wird, die nur additiv zur Erstausbildung hinzugefügt wird, ohne eine inhaltliche Verschränkung zu bewirken und somit keine qualitative Steigerung der Ausbildung erreicht. Hierdurch wird die Entwicklung der Lernenden im Hinblick auf eine Qualifikation auf Sacharbeiterebene mit europäischer Ausrichtung weitestgehend unterstützt, was mit dem Kontext der bildungspolitischen Zielrichtung übereinstimmt (vgl. ZB 1996/1997: 58).

K5 – Vollständige und faire Einschätzung: Im Hinblick auf das Ziel einer Integration eines Auslandspraktikums in den Bildungsgang kommt die WBL zu dem Schluss, dass eine solche systematische Integration nicht gelungen sei (vgl. AB 1999: 4). Auch andere 'kritische' Ergebnisse werden im Zusammenhang mit der Erhebung (Zielgruppenbezug) dargestellt: so zweifeln die potenziellen Abnehmer der Bildungsqualifikation an der Eignung der 'Berufsschule' als Kompetenzen vermittelnde Institution (vgl. AB 1999: 216). Darüber hinaus werden alle beabsichtigten Erkenntnisse dargestellt.

K6 – Offenlegung der Ergebnisse: Die für den MV Verantwortlichen werden genannt. Hierunter auch Ansprechpartner der Projektleitung und beteiligten Schulen sowie die Mitglieder der wissenschaftlichen Begleitung (ZB 1996/1997 und ZB 1997/1998: 4). Die Berichte sind über das MK des Landes Nordrhein-Westfalen zu beziehen.

K7 – Deklaration von Interessenkonflikten: Es werden keine Ursachen für mögliche Interessenkonflikte der an der Evaluation beteiligten Parteien genannt. Ex-ante formulierte schriftliche Vereinbarungen über Schutzmaßnahmen fehlen.

Resümee Standardbereich K:

Der durchschnittliche Ausprägungsgrad dieses Standardbereiches macht deutlich, dass für diesen Bereich ein Niveau erreicht wurde, das das Wohlergehen der Beteiligten und Betroffenen ausreichend berücksichtigt. Lediglich der Standard K7 (Deklaration von Interessenkonflikten), der vorbeugende Maßnahmen für mögliche Interessenkonflikte einfordert, ist von den Evaluatoren nicht erfüllt worden.

Vor allem aber die Hinweise, die die Berücksichtigung des Standard K1 (Unterstützung der Dienstleistungsorientierung) hervorheben und in die Berichte der Evaluation einfließen, tragen dazu bei, dass das Transferpotenzial des Modellversuchs auch in diesem Standardbereich gut ausgebildet wurde: Für Mitglieder der Bildungsplanung und Interessenten an der Überführung der Ergebnisse in einen Planungsmodellversuch wird die interne und externe Konsistenz des Lernfeldcurriculums hervorgehoben. Unterstützt durch die vollständige Offenlegung der Stärken und Schwächen des Modellversuchs (K6- Offenlegung der Ergebnisse) wird außerdem die Planungssicherheit für potenzielle Interessenten an den Evaluationsergebnissen (wissenschaftliche Ebene) erhöht.

G1 – MV-Dokumentation: Die Berichte geben Auskunft über den jeweiligen Stand des zu entwickelnden Curriculums (vgl. ZB 1996/1997: 95ff; ZB 1997/1998: 11ff; 27ff; AB 1999: 219ff). Durch den Vergleich der jeweiligen Curriculumfassungen können die Unterschiede, die als Ergebnisse des Entwicklungsprozesses anzusehen sind, erkannt werden. Die einzelnen Schritte des Entwicklungsprozesses sind dabei auf die interne Evaluation auf Schülerebene zurückzuführen, die im Frühjahr 1998 vorgenommen wurde, so dass die Curriculumfassungen im Zwischenbericht 1997/1998 und im Abschlussbericht 1999 die Konsequenzen aus den Erkenntnissen der Evaluation enthalten. Des Weiteren kann durch den Rückgriff auf die Informationen des Zwischenberichts 1996/1997 (Überblick über die Zusatzqualifikation 'Euro-Wirtschaftsassistent') der Stand der Ausgangsüberlegungen nachvollzogen werden, so dass die Berichte den vollständigen Entwicklungsprozess dokumentieren.

G2 – Kontextanalyse: Um bildungspolitische Entwicklungen einzubeziehen, sind regelmäßige Treffen der 'Dezernentengruppe' initiiert worden (vgl. ZB 1996/1997: 63). Mit in die Betrachtung sind außerdem bildungspolitische Prozesse, wie die laufende Überarbeitung bestehender Richtlinien (Groß- und Außenhandel) und der Wegfall der Fächerorientierung, eingeflossen (vgl. ZB 1996/1997: 62; vgl. AB 1999: 115; parallel laufende Aktivitäten). Für diesen Modellversuch außerdem relevant sind die aufgezeigten Aspekte der europäischen Dimension im Unterricht, die Bezugspunkte zur Entwicklung des Curriculums darstellen. Auch diese wurden von der WBL berücksichtigt (vgl. ZB 1996/1997: 58). Die Durchführungsorte werden charakterisiert, die beteiligten Personen genannt.

G3 – Beschreibung von Zielen und Vorgehen: Die WBL nennt die von ihnen verfolgten Ziele, differenziert nach Zielen des Modellversuchs und Zielen der Evaluation (vgl. ZB 1996/1997: 59 u. 65). Allerdings unterbleibt eine Operationalisierung der Evaluationsziele in dem Maße, dass nach Abschluss der Evaluation Aussagen zum Ausmaß der Zielerreichung möglich wären. Die zwei Evaluationsziele, die von der WBL als 'Forschungsfragen' bezeichnet werden, werden in Form von Fragen formuliert, so dass diese (nur) als 'bedingt' operationalisiert angesehen werden können (vgl. ZB 1996/1997: 65), was die Einordnung in den höchsten Ausprägungsgrad verhindert.

Das Vorgehen bei der Evaluation wird nicht differenziert nach formativer und summativer Evaluation; die auf beiden Ebenen zum Einsatz kommenden Instrumente werden in den Berichten beschrieben und deren Auswahl begründet (vgl. AB 1999: 13ff); außerdem wird deutlich, wie mit den Ergebnissen der summativen Evaluation weiter verfahren werden kann, da auf sie in den Abschlussempfehlungen der WBL Bezug genommen wird (vgl. AB 1999: 214ff).

G4 – Verlässliche Informationsquellen: Aus den Berichten geht die Anzahl der befragten Schüler hervor; deutlich wird jedoch nicht, nach welchen Kriterien diese Probanden ausgewählt wurden, so dass offen bleibt, ob und in welchem Maß die 'Quellen' der Evaluation beeinflusst waren. Auf Grund der allgemeinen Angaben zu den Schulstandorten des Modellversuchs kann davon ausgegangen werden, dass nicht alle Schüler der jeweiligen MV-Klassen in die Evaluation einbezogen wurden (vgl. ZB 1996/1997: 12ff u. AB 1999: 41ff), so dass sich die Erkenntnisse aus der Evaluation nicht auf die Grundgesamtheit der betroffenen Schüler beziehen. Die eingesetzten Fragebögen bzw. die grundlegenden Informationen über die Gestaltung der narrativen Interviews sind in den Berichten enthalten.

In Bezug auf die Auswahl der Interviewpartner (Probanden) aus der Berufspraxis (Personalleiter) erfolgte eine Samplebildung nach Region, Branche, Gesellschaftsform, Unternehmensgröße und internationale Markttätigkeit, so dass hier die Auswahlkriterien zu einer Steigerung der 'Verlässlichkeit der Informationsquellen' beitragen.

G5 – Valide Informationen: Die Evaluationsmethoden sind im Hinblick auf die formulierten Ziele geeignet, die gewünschten Informationen zu erheben, da die Inhaltsbereiche der Befragungen darauf abzielen zu prüfen, inwiefern Lernfelder als Ordnungsstrukturen für ein Curriculum geeignet sind, den inneren und äußeren Zusammenhang des Curriculums sicher zu stellen (siehe D1 - Praktische Verfahren). Der Einsatz der Evaluationsinstrumente wird ausführlich beschrieben und die Ergebnisse hieraus werden in Beziehung zueinander gesetzt (Kreuztabelle und Vergleich der Ergebnisse aus den Erhebungswellen). Dezidierte Aussagen über das Maß der Zielerreichung müssen jedoch unterbleiben, da die Ziele (siehe G3- Beschreibung von Zielen und Vorgehen) nicht vollständig operationalisiert wurden.

Im Hinblick auf die Auswahl der eingesetzten Evaluationsinstrumente sind die Informationen in den Evaluationsdokumentationen nur bedingt geeignet, diese zu begründen: Für die Auswahl resp. Gestaltung der Schüler- und Lehrer-Fragebögen werden keine Informationen gegeben; auch fehlen Hinweise auf bisherige Einsatzbereiche dieser Instrumente. Die 'qualitative Befragung' der Fremdsprachelehrer wird im Hinblick auf die von ihr zu erhebenden Daten (Informationen über die subjektive Interpretation des Lehrplans) legitimiert (vgl. AB 1999: 119). Durch den Bezug auf Aussagen aus der Fachliteratur werden die Ausführungen unterstützt.

In den Berichten sind jedoch keine Informationen über die Validität des hier spezifisch zum Einsatz kommenden selbst entwickelten Interviewtyps und der Fragebögen enthalten; auch fehlen Hinweise auf die Erprobung der Interviews/Fragebögen in anderen Bereichen. Die Auswahl der 'schwach strukturierten Interviews' mit den Vertretern der Unternehmen erfolgt durch den Hinweis, dass so die Gesprächsfüh-

rung flexibel bleibe und der Erfahrungsbereich der Befragten besser erkundet werden könne (vgl. Ab. 1999: 202).

G6 – Reliable Informationen: Die Berichte enthalten nur begrenzt Angaben über mögliche systematische Faktoren, die die Ergebnisse verfälschen könnten: Es erfolgt lediglich der Hinweis, dass der Fragebogen der zweiten Erhebungswelle in geringem Maße modifiziert werden musste, da die Anzahl der Lernfelder variierte und außerdem eine zusätzliche Frage den Bereich des Transfers der Ergebnisse abdecken sollte. Des Weiteren ist den Dokumentationen zu entnehmen, dass die Anzahl der ausgewerteten Fragebögen in den Erhebungswellen 'nahezu' identisch sei; die Anzahl variierte um einen Fragebogen (vgl. AB 1999: 41). Maßnahmen zur Reduzierung systematischer Fehler werden nicht ergriffen, auch werden die Nachteile der Erhebungsmethode der Befragung nicht durch andere Methoden ausgeglichen. Lediglich die Bandbreite der befragten Beteiligten und Betroffenen könnte als Ausgleichsversuch systematischer Fehler in den unterschiedlichen Probandengruppen angesehen werden; dass die Evaluatoren durch die breite Fächerung ihrer befragten Personengruppen dieses Ziel verfolgen, ist den Berichten jedoch nicht zu entnehmen.

G8 – Analyse quantitativer Informationen: Die Auswertungen der quantitativen Daten aus den Erhebungen mittels Fragebögen werden in Form von Tabellen dargestellt, wobei die Ergebnisse der zwei Erhebungswellen zwecks Vergleich gegenübergestellt werden. Den Dokumentationen ist zu entnehmen, auf welche jeweilige Grundgesamtheit die Aussagen bezogen sind und wie viele Fragebögen an den jeweiligen Schulstandorten ausgefüllt wurden. Die Qualität der Darstellung wird dabei durch das Gegenüberstellen von Ergebnissen, deren zu Grunde liegenden Fragen miteinander korrespondieren, in Form von Kreuztabellen unterstützt. Das Aufzeigen von absoluten und relativen Häufigkeiten rundet die Ergebnisdarstellung ab und trägt zu einer vertiefenden Interpretation der Daten bei, die insgesamt als nachvollziehbar bezeichnet werden kann (vgl. AB 1999: 41ff), so dass die Ausführungen die eingesetzten Evaluationsmethoden legitimieren.

G9 – Analyse qualitativer Informationen: Die Ergebnisse der 'qualitativen Interviews' (Fremdsprachlehrer; Grundgesamtheit: 8 Personen) wurden einer 'generalisierenden Analyse' unterzogen, d. h. Gemeinsamkeiten wurden von abweichenden Antworten getrennt und in die Bereiche 'Gemeinsamkeiten und Besondere' unterteilt, um einer künstlichen Homogenität zu entgehen. Dieses Vorgehen wurde für jede der elf Frage-Kategorien wiederholt. Die Interpretation der Ergebnisse erfolgt innerhalb dieser Kategorienbereiche, wobei die Schlussfolgerungen und Empfehlungen daraus sich auf die unsystematisierte Datenmenge beziehen (vgl. AB 1999:179ff).

Auch die Auswertung der Interviews mit den Unternehmensvertretern erfolgt kategoriengeleitet (sechs Kategorien) - und ebenfalls beziehen sich die 'zusammenfassenden Thesen' auf die unsystematisierte Datenmenge. Insgesamt ist ein Zusam-

menhang von Kategorien und Zielformulierungen der Evaluationsebene zu erkennen, so dass diese dazu beitragen, die gestellten 'Ziel-Fragen' zu beantworten. Eine Prüfung der Kategorien im Hinblick auf Validität und Reliabilität kann in beiden Fällen den Berichten nicht entnommen werden. Es erfolgt für die Auswertung der Interviews mit den Fremdsprachenlehrern lediglich der Hinweis, dass die Tonbänder ein 'zweites Mal abgehört' wurden, um 'Fehlinterpretationen' vorzubeugen (vgl. AB 1999: 139).

G10 – Begründete Schlussfolgerungen: Die Schlussfolgerungen aus den verschiedenen Teilen der Evaluation beziehen sich auf die Fragestellungen der Evaluationsebene, indem auf den inneren und äußeren Zusammenhalt von Lernfeldern für ein Curriculum abgestellt wird. Dabei wird jeweils der Bezug zu den verschiedenen Erhebungsebenen (Schüler, Lehrer, Unternehmensvertreter) genommen und auch die Interpretationen beziehen sich auf die jeweilige Ebene (jedoch ohne Kategorienbezug im Bereich der qualitativen Daten), so dass dieses Vorgehen als entwicklungslogisch bezeichnet werden kann, da sich auch die Verfahren der Evaluation in den Auswertungen widerspiegeln (vgl. AB 1999 ab Seite 41). Des Weiteren gehen auch kritische 'Stimmen' in die Auswertungen ein (vgl. AB 1999: 214ff), allerdings unterbleiben Rückschlüsse zur Modellversuchsebene, die sich in einer (hier nicht vorhandenen) zusammenfassenden Wertung des Modellversuch wieder finden.

G12 – Metaevaluation: Eine Metaevaluation ist nach Lage der Berichte nicht durchgeführt worden.

Resümee Standardbereich G:

Die Genauigkeitsstandards erreichen ein durchschnittliches Ausprägungsniveau, das zum Ausdruck bringt, dass die eingesetzten Evaluationsmethoden Informationen generieren, die den überwiegenden Teil der Fragestellungen auf Evaluationsebene beantworten können:

Die Modellversuchsdokumentationen, bestehend aus zwei Zwischen- und einem Abschlussbericht, geben Auskunft über die ursprünglichen Planungen zur Ausgestaltung des Modellversuchs und zeigen auch die Entwicklungsschritte im Zeitablauf auf, was in der Einstufung des Standards G1 in den höchsten Ausprägungsgrad zum Ausdruck kommt. Die hier schon vorhandene Transparenz wird durch die ausgeprägte Kontextanalyse (Standard G2) weiter ausgebaut. Nicht nur parallel laufende Aktivitäten werden aufgezeigt, sondern auch die Einordnung des Modellversuchsgegenstandes in bildungspolitische Entwicklungen wird dokumentiert.

Neben den nach Modellversuchsebene und Evaluationsebene differenzierten Zielen der WBL, werden auch die zum Einsatz kommenden Evaluationsinstrumente beschrieben (Standard G3 – Beschreibung von Zielen und Vorgehen), allerdings unterbleibt hier eine vertiefende Argumentation für die Auswahl des Instrumentariums. Leichte Lücken zeigen sich dann auch hinsichtlich der Erfüllung des Standards G4 - Verlässliche Informationsquellen, da die Auswahl der als Probanden bestimmten Schüler nicht erläutert wird, so dass das Maß der 'Beeinflussung' der 'Datenquellen' nicht nachvollzogen werden kann.

Um die Eignung der Evaluationsinstrumente als Mittel der Datenerhebung beurteilen zu können, ist es Aufgabe der Evaluatoren, Aussagen über deren Validität (G5 – Validität) und Reliabilität (G6 – Reliabilität) zu tätigen. Informationen hierzu sind in den Berichten jedoch nur sehr begrenzt enthalten. Positiv zeichnet sich dagegen die Auswertung quantitativer Daten (G8 – Analyse quantitativer Informationen) in der Gesamtbewertung dieses Standardbereiches ab, da diese nachvollziehbar geordnet und aufbereitet werden. Gleiches gilt für die Interpretationen, die hierauf aufbauend dokumentiert sind.

Leichte Schwächen zeigen sich in der Erfüllung der Standards G9 – Analyse qualitativer Informationen und G10 – Begründete Schlussfolgerungen, da sich hier die Empfehlungen nicht auf die zuvor gebildeten Kategorien beziehen, so dass Aussagen nur zur unsystematisierten Datenmenge in Bezug gesetzt werden können. Wäre eine Metaevaluation durchgeführt worden (G12 – Metaevaluation), hätten die beschriebenen Lücken auch für den G - Standardbereich aufgedeckt werden können, was zu einer Qualitätssteigerung insgesamt führen würde.

Das Kongruenzschema zum Nachweis der Korrespondenz von Evaluationszielen und Evaluationsinstrumenten kann vollständig angewendet werden:

Die Ziele der Evaluation werden in Form von zwei Fragen dokumentiert, die von den Evaluatoren als 'Forschungsfragen' bezeichnet werden, da sie die Evaluation als Beitrag zur 'Theoriebildung' ansehen (vgl. ZB 1996/1997: 64f). Fragen können als erster Operationalisierungsschritt angesehen werden und verlassen damit die Ebene der Formalziele. Allerdings unterlassen die Evaluatoren die weitere Operationalisierung, z. B. durch die Untersetzung mit Ausprägungsgraden, der Einsatz multipler Indikatoren unterbleibt also, so dass keine Aussagen über den Grad der Zielerreichung möglich sind.

Positiv ist festzustellen, dass die zum Einsatz kommenden Evaluationsinstrumente vollständig und nachvollziehbar dokumentiert und nach Einschätzung auf Grundlage der Standards D1 – Praktische Verfahren und G3 – Beschreibung von Zielen und Vorgehen auch geeignet sind, den Großteil der gewünschten Daten zu erheben. Da sich die Darstellung der Ergebnisse sich jedoch nicht auf zuvor gebildete Inhaltskategorien bezieht, nimmt die Qualität der Empfehlungen, die die Evaluatoren aussprechen, ab, was sich auch in der Gesamtwürdigung des betroffenen Standardbereiches ausdrückt.

Zwischenergebnis – Analyse und Interpretation der Ergebnisse

Die Auswertung der verschiedenen Erhebungen aus den Schülerbefragungen, den Lehrerinterviews und den Interviews mit den Vertretern der potenziellen Abnehmer der Auszubildenden bestätigen die festgelegten Lernfelder als Ordnungsmittel für das entwickelte Curriculum. Insofern kann der Modellversuch aus inhaltlicher Sicht als erfolgreich bezeichnet werden. Auch die getroffenen Aussagen und vorgenommenen Interpretationen sind, trotz mangelndem Bezug zu gebildeten Kategorien,

nachvollziehbar und mit Einschränkungen angemessen: Ein Ergebnis der Befragungen der Unternehmensvertreter ist, dass diese skeptisch den berufsbildenden Schulen als Ausbildungsinstitutionen für die Vermittlung der in kaufmännische Ausbildungsgänge integrierten Zusatzqualifikation gegenüber stehen. Die Unternehmensvertreter sprechen den Schulen und den Lehrern die didaktische und fachliche Kompetenz hierzu ab (vgl. AB 1999: 216). Hier hätte der Bezug zu den Auswertungskategorien (Marktmäßige Durchsetzung der Zusatzqualifikation, Ausbildungsbereitschaft u. a.) zu dem Ergebnis führen müssen, dass der Transfer resp. die feste Implementation des Curriculums in kodifizierter Form durch die mangelnde Bereitschaft der Unternehmen verhindert werde, was als Ergebnis im Abschlussbericht des Modellversuchs hätte hervorgehoben werden müssen.

Zur Einstufung der einzelnen Standards in den Standardbereichen ist festzustellen, dass vor allem im Bereich der G-Standards Schwächen auftraten, dieser Standardbereich jedoch in besonderem Maße für die Qualität der Aussagen zur Bewährung der Maßnahmen auf der Objektebene verantwortlich sind. Deutlich wirken sich diese Schwächen beim Standard G10 – Begründete Schlussfolgerungen aus (s. o.), so dass hierdurch der Modellcharakter dieses Erprobungsvorhabens insgesamt geschwächt wird.

6.7 Darstellung und Evaluation des Modellversuchs 'Ganzheitliche Lernansätze'

6.7.1 Beschreibung des Modellversuchs

6.7.1.1 Allgemeine und strukturelle Angaben

Der Modellversuch 'Entwicklung und Erprobung ganzheitlicher Lernansätze in der kaufmännischen Berufsausbildung am Beispiel des Ausbildungsberufs Versicherungskaufmann/-frau' wurde in der Zeit vom 01.09.1991 bis 31.12.1994 im Bundesland Saarland durchgeführt. Eine Projektleitung hatte die Koordination der Schulaktivitäten übernommen.

An dem Modellversuch beteiligt waren eine Klasse mit 20 Schülern und den darin eingesetzten fünf Lehrern (vgl. ZB 1992: Deckblatt⁴⁰). Darüber hinaus wurde eine Kontrollgruppe in die Auswertungen des Modellversuchs einbezogen (vgl. AB 1995: 7).

Die wissenschaftliche Begleitung wurde von der Universität des Saarlandes, Fachrichtung Erziehungswissenschaft, Fachgebiet Pädagogische Psychologie unter Leitung eines Universitätsprofessors wahrgenommen.

6.7.1.2 Modellversuchsziele und –inhalte/Tätigkeiten der WBL auf Modellversuchsebene

Das auf Wissensanhäufung ausgerichtete Curriculum im Ausbildungsberuf Versicherungskaufmann/-frau sollte durch eine Vernetzung der Inhalte geändert werden,

40) Der Zwischenbericht enthält keine Seitennummerierungen. Die angegebenen Seitenzahlen sind vom Verfasser hinzugefügt worden.

damit Schüler selbständiges, durch Hilfsmittel unterstütztes Arbeiten lernen konnten. Diese erarbeiteten unter Berücksichtigung der prüfungsrelevanten Stoffinhalte verschiedene Unterrichtsprojekte.

Hieraus ergaben sich Zielsetzungen, die sich auf verschiedene Bereiche und Ebenen in- und außerhalb des Modellversuchs beziehen⁴¹: Ermittlung lernpsychologischer Kriterien von Berufsschulunterricht, Erarbeitung von Merkmalen zur Gestaltung von Lehrplänen, Bestimmung von unterrichtsmethodischen Ansätzen, organisatorische Umgestaltungen, Entwicklung neuer Formen der Zusammenarbeit von Schule und Praxis, Überprüfung von Leistungsmessung und Leistungsbewertung, Auswirkungen auf die Lehrerfortbildung (vgl. AB 1995: 7f).

Die WBL hat im Rahmen ihres Selbstverständnisses eingangs eine Umfrage bei Lehrern, Ausbildern und Schülern durchgeführt, um subjektive Kriterien (Effizienzkriterien), die den Erfolg des Modellversuchs charakterisieren sollen, herauszuarbeiten; die Ergebnisse sind nach Angaben der WBL in den Modellversuchsprozess zurückgemeldet worden (vgl. ZB 1992: 12f, 18f).

Darüber hinaus nahm die WBL in regelmäßigen Abständen an den Teamsitzungen der Schulpraktiker teil.

6.7.1.3 Fortbildungen/Tagungen/modellversuchsbegleitende Transferaktivitäten

Es sind verschiedene Fortbildungsveranstaltungen zu den Themen der Themenzentrierten Interaktion, Initiierung von Gruppenprozessen, der Unternehmenskultur und der Rhetorik durchgeführt worden (vgl. AB 1995: 8).

6.7.2 Ergebnisse des Modellversuchs aus Sicht der WBL und Zwischenergebnis/Resümee

Dem Abschlussbericht ist unter dem Gliederungspunkte 'Ermittlung von lernpsychologischen Kriterien für den Berufsschulunterricht' zu entnehmen, dass zwei Rhetorikkurse und die Veranstaltungen zur Unternehmenskultur und zum Thema 'Gruppenprozesse' positive Ergebnisse hervorbrachten.

Darüber hinaus habe das Lehrerteam die geltenden Lehrpläne im Fach Wirtschafts- und Sozialkunde vernetzt und exemplarisch für eine Unterrichtsreihe mit einem didaktisch-methodischen Kommentar versehen. Außerdem seien in den Fächern Wirtschaftslehre, Sozialkunde, Deutsch, Buchführung und Religion 'alle' methodischen Ansätze ausprobiert worden; der Unterrichtsraum sei so gestaltet worden, dass die 'Bänke in U-Form' stünden. Im Hinblick auf die Zusammenarbeit von Schule und Praxis habe ein Erfahrungsaustausch mit den Ausbildungsbetrieben ergeben, dass diese an Veränderungen in der Schule interessiert sind (vgl. AB 1995: 10f).

41) Eine Trennung und nach Bereichen geordnete Darstellung der in den Dokumentationen angegebenen Ziele kann nicht erfolgen, da wegen der unübersichtlichen und lückenhaften Darstellung die Ausführungen zum Teil nicht nachvollziehbar sind.

Den Berichten sind keinerlei Informationen zu summativen Evaluationstätigkeiten zu entnehmen, entsprechend fehlen Informationen zu Ergebnissen hierzu, die der Metaevaluation unterzogen werden können.

Die dargestellten Ergebnisse auf Modellversuchsebene und die fehlenden Informationen zu einer summativen Evaluation durch die WBL nähren die Vermutung, dass die Mitglieder der wissenschaftlichen Begleitung einer Beratungs- und Steuerungsfunktion auf Modellversuchsebene sowie den Tätigkeiten auf Evaluationsebene nicht nachgekommen sind. Es liegen keine Modellversuchsergebnisse vor, die transferiert werden können, so dass der Modellcharakter dieses Erprobungsvorhabens verloren gegangen ist.

6.8 Datenauswertung und Interpretation

6.8.1 Zur Aussagefähigkeit und Darstellung der Mittelwerte der Standardbereiche

Die vorstehenden Ausführungen und Interpretationen zu den einzelnen Standards innerhalb der modellversuchsbezogenen metaevaluativen Dokumentenanalysen geben Auskunft über die Güte-Einschätzung der Konstruktionen zweiter Ordnung. Die Aussagekraft dieser Ausführungen resultiert dabei aus der Verbalisierung der auf die Evaluationsstandards bezogenen Stärken- und Schwächenanalyse der summativen Evaluation und ist wegen der Nachvollziehbarkeit und Transparenz des Vorgehens höher einzuschätzen, als es die Quantifizierung der Evaluationsstandards durch reines Nennen der Einordnung in einen absoluten Ausprägungsgrad zu leisten vermag. Um dennoch ein Gesamtbild zu erhalten, werden nachfolgend die ermittelten Durchschnittswerte der Standardbereiche dargestellt, da Quantifizierungen in einer Gesamtbetrachtung wiederum geeignet sind, Trends erkennen zu lassen. Die Mittelwerte aus der Metaevaluation der Nützlichkeitsstandards und Korrektheitsstandards sowie die der Durchführbarkeitsstandards und Genauigkeitsstandards werden dabei einerseits gegenüber gestellt ($N \leftrightarrow K$ und $D \leftrightarrow G$), um bestehende Abhängigkeiten aufzuzeigen, andererseits wird durch die partielle Auflösung der Standard-Grenzen (standardbereichübergreifende Darstellung) eine qualitative Steigerung der Aussagen angestrebt.⁴²

Gegenstand des Standardbereichs N ist das Identifizieren der an/von der Evaluation beteiligten und betroffenen Personen (Ermittlung des ‚Bedingungsfeldes‘) und der Orientierung der Evaluation an deren Informationsbedürfnissen; die K-Standards sollen darüber hinaus sicher stellen, dass die Evaluatoren die Auswirkungen des Modellversuchs im Hinblick auf die Bedürfnisse der Beteiligten und Betroffenen untersuchen und offen legen. D. h. durch den Standardbereich K werden, orientiert an den berücksichtigten Anforderungen der N-Standards, die Auswirkun-

42) Das Joint Committee on Standards for Educational Evaluation hebt hervor, dass jeder Standard dazu beiträgt, eines der Attribute Nützlichkeit, Durchführbarkeit, Korrektheit und Genauigkeit zu definieren, dass jedoch die Erfahrung gezeigt habe, dass ein Standard für mehr als ein Attribut von Bedeutung sein kann (vgl. Sanders 1999: 27). Daraus folgt, dass die Einordnung der einzelnen Standards durch das Joint Committee dort erfolgte, wo ein inhaltlicher Schwerpunkt zu erkennen ist.

gen der Evaluation auf die ermittelten Beteiligten und Betroffenen in den Mittelpunkt gerückt und somit ein direktes ergebnisbezogenes Abhängigkeitsverhältnis erzeugt.

Interdependenzen sind auch für den ausgewählten Durchführbarkeitsstandard und die Genauigkeitsstandards zu erkennen: Die zum Einsatz kommenden Evaluationsverfahren werden hinsichtlich ihrer Zweckrichtung und den erzeugten Störeinflüssen auf der Objektebene durch den Standard D1 – Praktische Verfahren bewertet; die Qualität der Evaluationsinstrumente und ihre Verwendung sowie Auswertung der hervorgebrachten Informationen findet wiederum ihren Ausdruck in den Genauigkeitsstandards. Während demnach die Durchführbarkeitsstandards den Fokus auf die Qualität der Evaluationsverfahren ausrichten, erweitern die Genauigkeitsstandards diesen Blickwinkel auf die Güte und/oder Verwendbarkeit der Informationen, die eine Evaluation (durch 'elaborierte Evaluationsverfahren'; siehe D1) über einen Modellversuch insgesamt generieren.

6.8.2 Interpretation der Mittelwerte der Standardbereiche

Im Bereich der Nützlichkeitsstandards erzielen die sechs untersuchten Modellversuche einen Mittelwert von 1,13 auf der von 0 bis 3 reichenden Ausprägungsgradskala. Hier erreicht dieser Standardbereich demnach ein Niveau, das zum Ausdruck bringt, dass die Evaluatoren die Informationsbedürfnisse der Beteiligten und Betroffenen nur zu einem geringen Teil berücksichtigen. Sollten die Erkenntnisse, die durch die Evaluation hervorgebracht werden, die Informationsbedürfnisse der Beteiligten und Betroffenen befriedigen, dann handelt es sich hierbei nicht um eine zielgerichtete Generierung von angeforderten Informationen. Die Gefahr, dass die gewonnenen Erkenntnisse ungenutzt bleiben, da diese für die B&B nicht von Bedeutung sind, ist groß.

N1	N3	N4	N5	N	K1	K5	K6	K7
1,16	0,83	1,5	1,0	1,13	0,83	0,5	2,5	0,0
G1	G2	G3	G4	G5	G6	G8	G9	G10
1,0	1,3	1,0	1,0	0,67	0,5	0,83	0,4	0,33

Tab. 10: Mittelwerte der Evaluationsstandards

Die mangelnde Orientierung an den Interessen der Beteiligten und Betroffenen zeigt sich auch in der durchschnittlichen Ausprägung des Standards K1: Hier ist es Aufgabe der Evaluatoren Ergebnisse zu präsentieren, die geeignet sind, als Entscheidungshilfe für nachfolgende Planungsmodellversuche oder als Grundlage für (berufs-) bildungspolitische Entscheidungen zu dienen. Wird dieser Evaluationsaufgabe nachgekommen, ist man der 'Dienstleistungsfunktion' des Standards gerecht geworden. Mit einem durchschnittlichen Ausprägungsgrad von 0,8 sind die Ergebnisse der Evaluation jedoch nur sehr begrenzt geeignet, diese Funktion zu erfüllen.

Zwei weitere Evaluationsstandards, die wechselseitig Einfluss nehmen, sind die Standards N5 – Klarheit des Berichts und K5 – Vollständige und faire Einschätzung. Der durchschnittliche Ausprägungsgrad des Standards N5 von '1' macht deutlich, dass Mängel hinsichtlich der Verständlichkeit und der Nachvollziehbarkeit der Ausführungen in den analysierten Modellversuchsdokumentationen bestehen, was u. U. auch auf die (nicht ausreichende) Berücksichtigung des Standards K5 zurückzuführen ist. Mit einem durchschnittlichen Ausprägungsgrad von 0,5 beim Standard K5 wird deutlich, dass i. d. R. nur beabsichtigte Erkenntnisse dokumentiert werden, so dass ungeplante Entwicklungen nicht aufgezeigt werden, die für das Verständnis des Modellversuchs (siehe Standard N5 und dessen Transfer) aber von Bedeutung sein können.

Die Korrektheitsstandards sind mit einem Mittelwert von 1,0 ebenfalls dem unteren Niveau der Ausprägungsgrade zuzurechnen. Zurückzuführen ist dies neben den genannten Faktoren vor allem auch auf Standard K7 (Deklaration von Interessenkonflikten), der in keinem der untersuchten Fälle erfüllt wurde. Ein abweichendes Ergebnis zeigt sich beim Standard K6 (Offenlegung der Ergebnisse); dieser bezieht sich auf die Veröffentlichungspflicht und stellt das Zugänglichmachen von Veröffentlichungen sowie das Nennen eines Ansprechpartners in den Vordergrund. Ein originärer Bezug zur inhaltlichen Qualität der Evaluation ist damit nicht zu erkennen, so dass die Erfüllung der Anforderungen dieses Standards (durchschnittlicher Ausprägungsgrad 2,5) von fünf der sechs Modellversuche nicht als Beitrag zur Qualitätssicherung der Evaluation im engeren Sinne interpretiert werden kann.

Die Durchführbarkeitsstandards erreichen einen durchschnittlichen Ausprägungsgrad von 1,2. Im Ergebnis sind hier drei von sechs Modellversuchen in den Ausprägungsgrad 0 einzustufen (Modalwert). Die Modellversuche EUWAS/NRW und SQ-Einzelhandel/Hamburg stehen dieser Einschätzung diametral gegenüber, sind also dem Ausprägungsgrad 3 zuzuordnen. Der Modellversuch SOL/Hessen erreicht in diesem Standardbereich den Ausprägungsgrad 1. Zurückzuführen ist das relativ schwache Ergebnis u. a. auf die in den Berichten lückenhaft beschriebenen Evaluationsverfahren bzw. auf die Schwächen oder das gänzliche Fehlen der Evaluationsinstrumente. D1 als Standard, der Aussagen zur Praktikabilität (Eignung zur zielgerichteten Informationssammlung und Störpotenzial auf der Objektebene) der eingesetzten Verfahren trifft, drückt durch den durchschnittlichen Ausprägungsgrad die nur bedingte Eignung der Instrumente und ein vertretbares Maß an Störungen auf der Objektebene aus.

Die Auswirkungen der Schwächen des Standards D1 zeigen sich auch bei den G-Standards - in einem geringen Niveau des durchschnittlichen Ausprägungsgrades. Mit 0,7 erreichen diese einen Mittelwert, der in etwa dem mittleren Bereich des Ausprägungsgrades 1 entspricht. – 'Die verwendeten Methoden der Evaluation sind lediglich geeignet, einen geringen Teil der Fragestellung zu erhellen und sind somit auch nur sehr bedingt zweckdienlich, die Güte und/oder Verwendbarkeit des zu evaluierenden Modellversuchs zu beurteilen' (siehe Evaluationsraster auf Basis ela-

borierter Evaluationsstandards; Anhang zu Teil B). Hervorzuheben sind dabei vor allem folgende Aspekte:

Eine Metaevaluation (G12) wurde in keinem der untersuchten Modellversuche durchgeführt. Die Anforderungen der Standards G1 (Programmdokumentation) und G2 (Kontextanalyse), die für das Verständnis der Modellversuchsentwicklungen und Rahmenbedingungen verantwortlich sind, konnten dagegen von der Mehrzahl (vier bzw. fünf) der Evaluationen zumindest ansatzweise erfüllt werden (Ø-Ausprägungsgrad: G1 – 1,0; G2 – 1,3). Alle weiteren Genauigkeitsstandards werden mit Ausnahme der in Hamburg und Nordrhein-Westfalen durchgeführten Modellversuche nicht bzw. nur schwer schwach berücksichtigt (Standards G3 bis G10).

Das schlechte Abschneiden der Evaluationen in diesem Standardbereich ist auf das Fehlen evaluationsbezogener Ziele, lückenhafter Ausführungen zur Validität und Reliabilität der Evaluationsinstrumente und Schwächen in der Auswertung, Darstellung und Interpretation der Evaluationsergebnisse zurückzuführen. Hier liegt die Vermutung nahe, dass die wissenschaftliche Begleitung sich auf die Unterstützung der Modellversuchspraktiker und die Entwicklung des Modellversuchsfeldes beschränkt und das Evaluieren ihres Vorgehens vernachlässigt hat.

Der Modellversuch ‚Euro-Wirtschaftsassistentin/Euro-Wirtschaftsassistent‘ (Nordrhein-Westfalen) erfüllt alle Standardbereiche im Verhältnis zu den anderen untersuchten Vorhaben überdurchschnittlich. Kritisch ist dennoch hervorzuheben, dass auf Hinweise zur Validität und Reliabilität der eingesetzten Evaluationsinstrumente weitestgehend verzichtet wurde. Somit sind Einschätzungen hinsichtlich der Eignung der eingesetzten Evaluationsinstrumente nur bedingt vornehmbar und dem Betrachter ist die Möglichkeit genommen, eine Entscheidung für die Übernahme des Explikationsmodells als Planungsmodellversuch vor dem Hintergrund der geforderten Informationen zu treffen. Ähnliche Aussagen finden für den Modellversuch ‚Schlüsselqualifikationen für Kaufleute im Einzelhandel‘, Hamburg, Anwendung. Zu bemängeln ist in diesem Modellversuch außer der für den MV EUWAS genannten Kritikpunkte auch die Auswertung der Evaluationsergebnisse. Hier leidet die Aussagekraft der Erkenntnisse zudem wegen fehlender Strukturierung der Ergebnisse.

Besonderes Augenmerk verdienen jedoch nicht nur die zwei Modellversuche, die relativ hohe Ausprägungsgrade innerhalb der untersuchten Standards aufweisen, sondern auch zwei Modellversuche, die nahezu keinen der Standards auch nur ansatzweise erfüllen konnten.

Ein Modellversuch (Saarland)⁴³ enthält den Hinweis auf eklatante Schwierigkeiten im Modellversuchsfeld, die die Zielerreichung insgesamt in Frage stellen (ZB

43) MV-Titel: ‚Entwicklung und Erprobung ganzheitlicher Lernansätze in der kaufmännischen Berufsausbildung am Beispiel des Ausbildungsberufes Versicherungskaufmann/-frau‘.

Gld. Pkt. 2.2). Gleichwohl empfiehlt die wissenschaftliche Begleitung die Fortführung des Vorhabens. Eine Begründung für diese Entscheidung unterbleibt. Eine summative Evaluation, die die Anwendung des Meta-Evaluationsrasters ermöglichen würde, ist nach Lage der Berichte nicht durchgeführt worden.

Auffällig ist außerdem der Modellversuch 'KUNO - Kundenorientierung in der kaufmännischen Berufsausbildung' des Bundeslandes Bremen. Die MV-Beschreibung ist wenig detailliert und die Berichte enthalten im Wesentlichen die Unterlagen, die im Rahmen des Erprobungsvorhabens erstellt wurden. Summative Evaluationstätigkeiten der WBL sind den Berichten nur bedingt zu entnehmen; die Ergebnisse sind darüber hinaus nicht geeignet, die Fragestellungen des Modellversuchs zu beantworten.

7 Zusammenfassung und Ausblick

7.1 Auswertung der deskriptiv-analytischen sowie der evaluativen Aspekte der Studie

Durch die 1971 geschaffene Rechtsgrundlage für das System der Modellversuchsförderung (vgl. Knauss 1978: 221ff.), ist ein Instrument installiert worden, das der Politikberatung dienen und darüber hinaus einen Beitrag zur Bildungsplanung leisten soll, um so Veränderungen im Bildungsbereich zu initiieren. Neben dieser originär politischen Dimension werden von Modellversuchen vor allem auch Impulse für die (Berufs-) Bildungsforschung erwartet, die sich aus der wissenschaftlichen Begleitung der Erprobungsvorhaben ergeben sollen. Die Vielzahl der intendierten Bereiche, für die das Instrument Ergebnisse resp. Erfolge hervorbringen soll, nährt die These, dass Modellversuche die Ansprüche im Hinblick auf Innovationskraft und Veränderungspotenzial für das Praxisfeld nicht erfüllen können.

Da die 'Leistungsfähigkeit' des Instruments Modellversuch nicht direkt ermittelt werden kann (siehe Kap. 1.2), ist es sinnvoll, die oben dargestellte Funktions-Trilogie aufzuspalten. Die Untersuchung der politischen Dimension erfolgt durch Explizierung und Operationalisierung der berufsbildungspolitischen Ziele der jeweiligen Bundesminister und Anlehnung dieser an ex-post Dokumentationen (Berufsbildungsberichte). Grundlage für dieses Vorgehen ist die Hypothese, dass Modellversuche keine 'pädagogischen Experimente' zur Verbesserung des Bildungswesens darstellen, d. h. in den berufsbildungspolitischen Zielintentionen keine Kohärenz zur Modellversuchsinitiierung wiederzufinden ist (Indiz 1).

Die Analyse der finanziellen Förderungsentwicklung im Rahmen der Untersuchung der politischen Dimension von Modellversuchen geht von der Hypothese aus, dass diese mittelbare Indikatoren darstellen, die Auskunft über die Bedeutung des Instruments in der Berufsbildungspolitik geben können (Indiz 2). Auch wenn kein direkter Zusammenhang von finanzieller Förderungshöhe und Outputniveau konstatiert werden kann (s. u.), so können doch Hinweise ermittelt werden, die eine mögliche Prioritätenverlagerung innerhalb der Berufsbildungspolitik vermuten lassen.

Der forschungsbezogenen Dimension (Berufsbildungsforschung) wird auf Grundlage der Hypothese nachgegangen, dass die Überprüfung der hervorgetretenen Modellversuchseffekte im Sinne einer Zielevaulation durch die wissenschaftliche Begleitung unterbleibt. Dabei wird die wissenschaftliche Begleitung und ihre Evaluationsaufgabe als integraler Bestandteil eines Modellversuches verstanden, so dass eine Korrelation von Evaluationsqualität und Innovationsgehalt angenommen werden kann, die auf Grundlage der Erkenntnisse auf der metaevaluativen Ebene, Schlussfolgerungen für die Objekt- oder Erprobungsebene zulassen (Leistungsmessung).

Mit Hilfe der gesammelten Indizien aus dem deskriptiv-analytischen Teil und in Kombination mit den Erkenntnissen aus dem metaevaluativen Teil der Studie, sind

nun Rückschlüsse möglich, die Aussagen über die Leistungsfähigkeit von Modellversuchen mit gleicher oder ähnlicher Versuchszintention wie die untersuchten Erprobungsvorhaben zulassen.

Um die Ergebnisse der Untersuchung strukturiert und nachvollziehbar darstellen zu können, wird der Aufbau der Erarbeitung auch für die nachfolgende Ergebnisfokussierung aufgegriffen (siehe 7.1.1 und 7.1.2).

Demzufolge werden in Kapitel 7.1.1 Ergebnisse und Schlussfolgerungen aus dem in Teil A vorgenommenen berufsbildungspolitischen Rückblick, im Sinne eines Vergleichs von Berufsbildungszielen und deren Erreichung in chronologischer Reihenfolge und Unterteilung in die Amtszeiten der jeweiligen Bundesbildungsminister, aufgezeigt. Auch werden die Erkenntnisse, die sich aus der Untersuchung der finanziellen Förderungsentwicklung ergaben, hervorgehoben.

Teil B beinhaltet die Entwicklung eines ex-post anzuwendenden metaevaluativen Rasters auf Basis der vom Joint Committee for Educational Evaluation vorgeschlagenen Evaluationsstandards. Dieses Raster wurde auf die in Kapitel 3.1.1 gruppierten Modellversuche mit Konzentration auf die Gruppe vier ‚Modellversuche zu Innovationen in bestehenden Fächern oder Schulformen‘ und hier auf die Modellversuche zum Themenbereich ‚Lehr-Lern-Methoden‘ angewandt. Die Resultate hieraus werden noch einmal aufgegriffen, um die aufgewiesenen Schwachstellen hervorzuheben.

Da sich die Erkenntnisse aus diesem Teil der Studie auf den Untersuchungszeitraum 1983 bis 1998 beziehen, ist, um die Aktualität der Studie zu wahren, eine Gegenüberstellung der Änderungen der BLK-Förderrichtlinien von 1998 mit den Kritikpunkten aus der metaevaluativen Untersuchung erforderlich (siehe hierzu 7.1.2). Im Ergebnis kann festgestellt werden, ob sich durch die Änderungen die Rahmenbedingungen für Modellversuche derart gewandelt haben, dass die aufgewiesenen Schwachstellen ‚geschlossen‘ sind und eine Steigerung der Leistungen dieses Instruments zu erwarten ist.

Abschließend werden im Gliederungspunkt 7.2 Vorschläge mit Diskussionscharakter dargestellt, die einer Verbesserung der Modellversuchsarbeit und der Nachhaltigkeit des Instruments dienen sollen.

7.1.1 Auswertung und Fazit des deskriptiv-analytischen Teils der Studie (Teil A)

- Ergebnisvorschau -

Es sind keine operationalen Zielformulierungen der verfolgten berufsbildungspolitischen Intentionen durch die jeweiligen Bundesminister festzustellen,

- d. h.**
- Aussagen über den Zielerreichungsgrad sind nicht möglich.
 - eine prospektive Initiierung berufsbildungspolitischer Entwicklungen wird verhindert.

Eine Kohärenz von Modellversuchsintentionen und berufsbildungspolitischen Absichten sind nur für einen begrenzten Zeitraum erkennbar,

d. h. Berufsbildungspolitik kann als re-agierendes Feld beschrieben werden, in dem Modellversuche nicht zwingend als Erprobungsmöglichkeit angesehen werden.

Trotz steigender Nettoausgaben des Bundes für Bildung und Wissenschaft ist eine negative Entwicklung der finanziellen Fördergrößen für das Instrument Modellversuch nachweisbar,

d. h. die Bedeutung, die man Modellversuchen als berufsbildungspolitischem Instrument beimisst, hat abgenommen.

Der bildungspolitische Rückblick hat gezeigt, dass die Bundesbildungsminister im Untersuchungszeitraum eine operationale Zielformulierung der von ihnen verfolgten berufsbildungspolitischen Intentionen unterlassen haben, so dass Aussagen zum Zielerreichungsgrad nicht möglich waren. Aus diesem Untersuchungsschritt geht vor allem der Zweifel an der Planmäßigkeit ihrer Handlungen hervor, die hierdurch insgesamt in Frage gestellt ist. Durch dieses nicht-agieren wird, daraus folgernd, eine prospektive Initiierung berufsbildungspolitischer Entwicklungen verhindert, so dass einerseits die Gefahr des uneffektiven Einsatzes (knapper) finanzieller Ressourcen (s. u.) groß ist und andererseits arbeitsmarktpolitische Entwicklungen oder gesicherte Erkenntnisse aus den relevanten Wissenschaftsgebieten nicht zielgerichtet aufgegriffen und in Modellversuchen umgesetzt werden können.

Zurückgeführt werden könnte dies auf den Kerngedanken der Ausführungen von Rolff, der besagt, dass Bildungsplanung lediglich auf den Vollzug bereits abgeschlossener Willensprozesse beschränkt ist, die das Resultat längst erfolgter sozialer Prozesse und politischer Entscheidungen sind (vgl. Rolff 1971: 319). Berücksichtigt man darüber hinaus den in Kapitel 3.3.1 aufgezeigten Wandel in der Berufsbildungspolitik zwischen Willms/Möllemann und Ortleb (Laermann)/Rüttgers, dann scheint vor allem das Bildungsverhalten der Jugendlichen als Nachfrager, Merkmal des angesprochenen sozialen Wandels zu sein, das starken Einfluss auf das bildungspolitische Gerieren der Minister hat und (auch hier) nicht das zielgerichtete und durch Modellversuchsergebnisse gestützte Handeln der verantwortlichen Politiker. In Zusammenhang mit bildungsökonomischen Planungsansätzen des Bildungs- und Arbeitskräftebedarfs lässt sich dieses Verhalten in den Rahmen angebotsorientierter Modelle einordnen, bei denen das Agieren der Arbeitskräfteanbieter, hier der Jugendlichen, im Mittelpunkt der Betrachtung steht (vgl. Huiskens: 1973: 146).

Darüber hinaus wurden die in 3.1.2 gruppierten Modellversuche den für den Untersuchungszweck operationalisierten berufsbildungspolitischen Zielen gegenübergestellt, um die zu erwartende Übereinstimmung von Berufsbildungspolitik und Modellversuchsinitiierung zu überprüfen. Hierbei ergab sich, dass bis 1991 eine Kohärenz von Modellversuchsinhalten und berufsbildungspolitischen Intentionen zu

konstatieren ist. Für den nachfolgenden Zeitraum kann diese Feststellung nicht getroffen werden.

Aufgrund dieser Erkenntnisse lässt sich Berufsbildungspolitik im Untersuchungszeitraum als ein eher 're-agierendes', denn als 'agierendes' politisches Feld beschreiben, in dem Modellversuche nicht zwingend als Erprobungsmöglichkeit angesehen werden, um mit geringerem Risiko Neuerungen zu adaptieren.

Die Aussagefähigkeit dieser Untersuchungsschritte insgesamt ist dahingehend angreifbar, als dass die vorgenommenen Zieloperationalisierungen auf der subjektiven Einschätzung des Autors beruhen. Sollte dies im Einzelfall zu Ungenauigkeiten geführt haben, dann könnte es zu Fehlinterpretationen gekommen sein. Jedoch kann mit dem Nichtvorhandensein geeigneter Zielformulierungen der Bundesminister die Notwendigkeit des gewählten Vorgehens begründet werden, so dass bei Ungenauigkeiten im Einzelfall das Vorgehen insgesamt nicht desavouiert werden kann. Die gleichen Einschränkungen gelten für die Aussagen, die sich aus der Untersuchung der Kohärenz von Modellversuchsinitiierung und Zielen in der Berufsbildungspolitik ergeben.

Ein weiteres Indiz, das an der Bedeutung des Instruments 'Modellversuch' für die Berufsbildungspolitik Zweifel aufkommen lässt, ist die Entwicklung der finanziellen Fördergrößen für BLK-Modellversuche (siehe 3.4). Die hierfür benötigten Daten wurden weitgehend aus den Angaben der statistischen Jahrbücher für die Bundesrepublik Deutschland der jeweiligen Bezugsjahre ermittelt. Auch hier sind marginale Ermittlungs- oder Rundungsfehler denkbar, deren Bedeutung sich im Zusammenhang und vor dem Hintergrund von beabsichtigten Trendaussagen relativieren.

Die Interpretation der aufgezeigten Größen im Hinblick auf einen direkten Input-Outputzusammenhang ist nicht möglich (siehe 3.4.2), weil davon auszugehen ist, dass ein steigender Mittelzufluss nicht zwingend an die Stellen im System 'Modellversuch' fließt, die direkten Einfluss auf das Ergebnisniveau haben. Eine Aussage in dem Sinne, dass zunehmende Fördermittel zu einer Qualitätssteigerung der Leistungen von Modellversuchen führen, ist demnach nicht zulässig. Konträr dazu sind jedoch zwei Möglichkeiten denkbar, die sinkende Fördermittel begründen können: Einerseits kann das finanzielle Gesamtvolumen für Bildungsausgaben abnehmen, so dass anteilmäßig auch die Ausgaben für Modellversuche geringer werden, andererseits kann auch die Bedeutung des Instruments innerhalb der berufsbildungspolitischen Prioritätenliste abgenommen haben, so dass weniger Mittel aufgewendet werden sollen.

Betrachtet man jedoch die in der Tendenz steigenden Nettoausgaben des Bundes für Bildung und Wissenschaft, in denen auch die Mittel für Modellversuche enthalten sind und richtet dabei das Augenmerk auf die negative Entwicklung des Anteils für Modellversuche an dieser Größe, dann ist offensichtlich, dass Modellversuche nicht mehr die (berufs-) bildungspolitische Bedeutung innehaben, die ihnen 1971 mit Aufnahme der Tätigkeiten durch die BLK zugewiesen wurde.

Aufgrund der vorstehenden Erkenntnisse aus dem deskriptiv-analytischen Teil (A) der Studie kann festgehalten werden, dass sowohl die gewonnenen Aspekte aus der berufsbildungspolitischen Analyse, als auch die Erkenntnisse aus der Untersuchung der finanziellen Förderungsentwicklung Indizien hervorbringen, die an der durch die BLK hervorgehobenen Bedeutung des Instruments ‚Modellversuch‘ für die Berufsbildungspolitik zweifeln lassen.

7.1.2 Auswertung und Fazit des evaluativen Teils der Studie (Teil B)

- Ergebnisvorschau -

Der Rücklauf der angeforderten Modellversuchsdokumentationen wies Schwächen auf,

d. h. es zeigte sich eine Bandbreite von Möglichkeiten den schwachen Rücklauf betreffend, die von der Nichtzusendung bzw. Nichterstellung der geforderten Unterlagen bis hin zur unterlassenen Evaluation reichen. Von den angeforderten Modellversuchsdokumentationen der 19 ausgewählten Modellversuche konnten daher letztendlich sechs für die Metaevaluation herangezogen werden.

Lediglich zwei der ausgewählten Modellversuche verfügen über evaluationsbezogene Ziele,

d. h. die Schwäche bei denjenigen Evaluationen, die darauf verzichtet haben, entsprechende Ziele zu formulieren trägt dazu bei, dass Ergebnisse nicht oder nur unter großer Unsicherheit transferiert werden können und der Modellcharakter verloren geht.

Der Ausweis aufgetretener Interessenkonflikte bzw. anderer Schwierigkeiten im Modellversuchsfeld unterbleibt,

d. h. auch hier wird der erfolgsgerichtete Transfer verzögert bzw. verhindert, da beim Übergang vom Explikationsmodellversuch zum Planungsmodellversuch potenzielle Probleme nicht im Voraus vermieden werden können.

Die eingesetzten Untersuchungs-/Evaluationsdesigns werden nicht dokumentiert,

d. h. in Verbindung mit den fehlenden evaluationsbezogenen Zielen, kann vermutet werden, dass die Mitglieder der wissenschaftlichen Begleitung ihre Tätigkeiten auf die Modellversuchsebene beschränkt haben bzw. es sind im Einzelfall die Tätigkeiten der WBL in Gänze nicht nachvollziehbar.

Der Standardbereich 'Genauigkeit' (Genauigkeitsstandards) wird zum Großteil nicht erfüllt,

d. h. die Einhaltung der sich auf die inhaltliche Qualität der Evaluation beziehenden Standards kann durch die untersuchten Modellversuche nicht ausreichend nachgewiesen werden.

Neben den Problemen beim Bezug der Modellversuchsdokumentationen (siehe Kap. 6.1) ist auf Ebene der Metaevaluationsergebnisse die Tatsache beachtens-

wert, d. h. im negativen Sinne einzuschätzen, dass lediglich die Modellversuche 'SOL-Hessen' und 'EUWAS-Nordrhein-Westfalen' über evaluationsbezogene Ziele verfügen, die zudem noch ausreichend operationalisiert und überprüft wurden, so dass auch eine Kongruenzprüfung von evaluationsbezogenen Zielen und eingesetzten Evaluationsverfahren möglich war und somit Aussagen zum inhaltlichen Erfolg der Modellversuche sowie eine Einschätzung der Interpretationen und Aussagen möglich sind. Alle anderen untersuchten Modellversuche haben auf diese wesentliche Forderung verzichtet und ihre Tätigkeit auf das Modellversuchsfeld konzentriert. Eine distanzierte Reflexion, wie sie in traditionellen Forschungsdesigns gefordert wird oder aber eine Wissenschafts-Praxis-Kommunikation mit dem Ziel des Austausches unterschiedlicher Perspektiven und Interessenschwerpunkte (vgl. Euler 1993: 204; Euler u. a.: 1988: 2; Zimmer 1997: 28; Euler, Sloane: 1998: 312f), als Merkmale neuerer Evaluationsansätze, ist hier nicht aufzufinden. Im Rahmen dieser Konzentration auf das Modellversuchsfeld ist zudem auch der Ausweis aufgetretener Interessenkonflikte bzw. Schwierigkeiten bei der Durchführung der Modellversuchstätigkeiten unterblieben.

Ein weiterer Kritikpunkt bezieht sich auf die fehlende Beschreibung der eingesetzten Untersuchungs-/Evaluationsdesigns. In Zusammenhang mit den nicht vorhandenen Evaluationszielen kann vermutet werden, dass die Mitglieder der wissenschaftlichen Begleitungen ihr Aufgabenverständnis an der Entwicklung des Modellversuchsfeldes ausgerichtet haben. Im Extremfall (siehe MV KUNO/Bremen oder MV ganzheitliches Lernen/Saarland) kann sogar gemutmaßt werden, dass zumindest phasenweise, wenn nicht vollständig eine wissenschaftliche Begleitung nur formal bestanden und diese ihre Tätigkeiten nur zu Dokumentationszwecken aufgenommen hat.

Die genannten Problembereiche wirken sich insbesondere auf den Bereich der Genauigkeitsstandards aus, die sich auf die inhaltliche Qualität der Evaluation (Evaluation im engeren Sinne) beziehen. Hier sind es vor allem die Standards Validität und Reliabilität sowie die Analyse quantitativer und qualitativer Informationen, die nicht bzw. nur von zwei Modellversuchen (wenn auch nicht vollständig) erfüllt wurden.

Die vorstehenden Ausführungen weisen auf Schwächen in der Ausgestaltung der wissenschaftlichen Begleitung resp. der Evaluation hin. Dehnbostel und Euler, die sich im Rahmen eines Workshops mit diesem Problem beschäftigt haben, konstatieren: Von der Legitimation einer Forschungsmethode über das Erkenntnisproblem "[...] losgelöst ist die Frage, mit welcher Kompetenz und mit welchem Grad an Phantasie eine Forschungsmethode im Rahmen der Erkenntnisgewinnung nicht nur eingesetzt, sondern umgesetzt bzw. ausgeprägt wird. In diesem Zusammenhang (gemeint ist die Diskussion um die Art der einzusetzenden Forschungsmethode; Anm. Verf.) wird gelegentlich eingewendet, dass zum einen den qualitativ ansetzenden Methoden noch die Bindung an überzeugende Forschungsstandards fehle, zum anderen die Breite der eingesetzten Methoden den Verdacht eines Dilet-

tantismus hinsichtlich der kompetenten Anwendung dieser Methoden nähre" (Euler/Dehnbostel 1998: 497).

Ungeachtet begrifflicher Ungenauigkeiten, die sich m. E. im vorstehenden Zitat aus der Gleichsetzung von 'Evaluationsmethoden mit Forschungsmethoden' ergeben und die auch eine vom Zweck der Evaluation abweichende Intention bei Evaluationsvorhaben erwarten lässt, kann die hierin enthaltene Kritik an der Evaluationspraxis, um die Ausführungen und Erkenntnisse zur 'Evaluationsforschung', siehe Kapitel 4, erweitert werden: es ist zweifelhaft, ob überhaupt realisierbare Evaluationsansätze eingesetzt resp. genutzt werden und ob darüber hinaus Modellversuche derart gestaltet werden, dass die Evaluatoren auf eine 'theoretischen Fundierung' zurückgreifen können. Die Ergebnisse der metaevaluativen Untersuchung zeigen Indizien auf (z. B. Standards N5, K5, K6, K7, G3), die die These stützen, dass eine derart gestaltete Evaluation mit einer rationalen Ziel- und Maßnahmenplanung (siehe auch Kap. 4) in Modellversuchen nicht zum Tragen kommt. Im Ergebnis handelt es sich demnach um Evaluationen, die nicht auf erklärende oder vermittelnde Variablen aufbauen (siehe Kap. 4.2ff), und somit als Black-Box-Modelle der Evaluation bezeichnet werden und keine Erklärung bieten können, wie ein Programm, ein Modellversuch funktioniert, d. h. wie die intendierten und nicht-intendierten Effekte entstehen.

Fokussierend auf die festgestellte Kohärenz von Evaluationsleistung und Innovationsgehalt für das Praxisfeld (siehe Kap. 1.2) sowie unter Hinzuziehung der Indizien aus dem deskriptiv-analytischen Teil der Studie lassen die Ausführungen den Schluss in Bezug auf die 'Leistungsfähigkeit' des bildungspolitischen Instruments zu, dass Modellversuche, die eine ähnliche Struktur aufweisen wie die untersuchten Erprobungsvorhaben keine wichtigen Entscheidungshilfen für die Entwicklung des Bildungswesens geben, wie es von der Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung erwartet wird (BLK 1998: 7). Selbst die Einsicht, "[...] dass Modellversuche in aller Regel weit weniger spektakuläre Ergebnisse erbringen, als dies ursprünglich wohl von den Beteiligten und der Öffentlichkeit erwartet worden ist" (BLK 1981: 6), scheint zu optimistisch.

Ein möglicher Schwachpunkt der Herangehensweise im metaevaluativen Teil dieser Studie ergibt sich aus der Auswahl (Bestimmung der anzuwendenden Evaluationsstandards per Ausschlussverfahren) der hinzugezogenen Evaluationsstandards sowie aus der Einschätzung der Merkmalsausprägungen der untersuchten Modellversuche durch den Autor. Dieses Problem hätte durch eine noch ausführlichere Argumentation bei der Einstufung der Ausprägungsgrade der untersuchten Standards gemildert werden, die Subjektivität aber nicht vollständig eliminieren können. In Analogie zu Stufflebeam gilt gleiches für seine Vorgehensweise, die hier hätte verwendet werden können; dieser hatte bei der Nicht-Erfüllung der von ihm als zwingend erachteten Evaluationsstandards das zu untersuchende Evaluationsmodell insgesamt als 'nicht-geeignet' eingeschätzt (siehe Stufflebeam 2001). Dass allerdings die inhaltliche Gestaltung und die Außenwirkung von BLK-Modellversuchsprogrammen verbesserungswürdig und somit Kritik grundsätzlich

angemessen und unerlässlich ist, zeigt die Neuordnung der Modellversuchsförderung. Westkamp hebt die Notwendigkeit von Änderungen hervor und bezieht sich dabei auf geäußerte Kritik an Modellversuchen im Bereich der allgemeinen Bildung und verweist vor allem auf die geringe Anzahl transferierter Modelle (vgl. Westkamp 2000: 291). Dieser Aspekt nährt die Vermutung, dass auch hier Schwächen in der wissenschaftlichen Begleitung und Evaluation die Übernahme von Modellversuchsergebnissen behindern.

Ein weiterer Aspekt, der die Reichweite der Aussagen einschränken könnte, bezieht sich auf die zeitliche Dimension der durchgeführten Untersuchung. Durch die Änderungen bei der Förderung von Modellversuchen durch die BLK, die am 01.01.1998 in Kraft getreten sind (BLK 1998), bestünde die Möglichkeit, dass die oben aufgezeigten Problemkreise erkannt und durch neue Anforderungen vermieden werden. Daher ist es notwendig, die wesentlichen Änderungsgedanken und -aspekte, die den Anstoß zur Neuausrichtung der Modellversuchsförderung gaben nachfolgend aufzuzeigen und der in der Studie ermittelten wesentlichen Schwachstelle (wissenschaftliche Begleitung/ Evaluation) gegenüberzustellen.

Ziel der Neukonzeptualisierung war die Fokussierung der Modellversuchsförderung auf wenige 'politisch verwertbare Programmthemen', die in einem Konzept der 'Programmförderung' die bisherige Einzelförderung ablösen sollte (vgl. Westkamp 2000: 291). Dabei werde nicht der Anspruch erhoben, das gesamte Bildungswesen weiterzuentwickeln, vielmehr werde angestrebt, "[...] im politischen Konsens mit Bund und Ländern sich auf notwendige Themen zu einigen, die vordringlich und wirkungsvoll in die Breite umzusetzen sind" (Westkamp 2000: 292). Die daraufhin beschlossenen Änderungen beziehen sich auf die Aspekte Zeitdauer eines Programms, Finanzvolumen, Entscheidungskompetenzen, Vorschlagsrecht und Ausarbeitung eines Programmkonzepts, Durchführung durch Programmträger sowie Finanzierung und Abrechnung (vgl. BLK: 1998, Westkamp 2000: 293f).

Der Aspekt der Evaluation, der als wesentliche Schwachstelle von Modellversuchen in dieser Arbeit herausgestellt wurde, wird im Konzept der Programmförderung dagegen nur insofern angesprochen, als dass eine zeitliche Begrenzung der Fördermaßnahmen notwendig ist, damit diese zeitnah evaluiert werden können. Über die Qualität oder die Ausgestaltung von Evaluationen gibt es keinerlei Hinweise, so dass aus der Änderung der Modellversuchsförderung durch die BLK keine qualitative Änderung der wissenschaftlichen Begleitung und deren Evaluation zu erwarten ist und die Erkenntnisse aus der vorliegenden Studie nicht nur ex-post, sondern auch für zukünftige Modellversuche relevant sind.

7.2 Diskussionsansätze zur Steigerung der Leistungsfähigkeit des Instruments 'Modellversuch'

Mögliche Änderungen, die zur Stärkung der Leistungsfähigkeit des Instruments Modellversuch beitragen können, sollten auf zwei Ebenen diskutiert werden: Gemeint sind die berufsbildungspolitische (ggf. administrative) und die modellversuchs- bzw. evaluationsbezogene Ebene.

Im Hinblick auf die Berufsbildungspolitik geht es um die Problematik der Modellversuchsinitiierung, genauer um die Frage, ob Modellversuche in den Kontext der bildungspolitischen Intentionen als innovatives ‚Anschubelement‘ einbezogen werden sollen oder ob der umgekehrte Weg im Sinne eines ‚bottom-up‘-Prinzips stringentes Verfahren wird.

Wegen der festgestellten geringen Kohärenz von berufsbildungspolitischen Zielen und Modellversuchsinhalten und der wenig flexiblen politischen Reaktionsmöglichkeit auf berufsfeldbezogene Neuerungen, ist das zweite Verfahren, das Entwicklungen aus dem ökonomischen, technischen oder pädagogischen Bezugssystem (Berufsbildungspraxis im weitesten Sinne und Berufs- und Wirtschaftspädagogik) aufgreift zu begrüßen, um auf diesem Wege die Aktualität der Berufsbildungspraxis zu gewährleisten. In diesem Zusammenhang bietet die Programmförderung, wie sie seit 1998 von der BLK durchgeführt wird, Orientierungspunkte politisch verwertbarer und am Bezugssystem ausgerichteter Programmenthemen und wirkt somit einer möglicherweise vorhandenen Beliebigkeit beantragter Modellversuchsinhalte entgegen. Insgesamt wird somit der Aspekt der bildungspolitischen Planung und der sinnvollen Ausgestaltung der Bandbreite eines Förderprogramms vereinigt.

Um jedoch die von der BLK vorgesehene Erfolgsorientierung durchzusetzen, die ihren Ausdruck u. a. in der Forderung nach "[...] rascher Verfügbarkeit und Umsetzbarkeit der Ergebnisse innovativer Entwicklungen [...]" (BLK 1998: 2) findet, ist eine präzisere Definition der Aufgaben der wissenschaftlichen Begleitung durch die BLK erforderlich. Denkbar ist eine Bandbreite differenter Begleit-Konzepte, wie sie sich beispielhaft auch in den Anträgen zum Modellversuchsprogramm 'innovelle-bs' (Laufzeit 02/2001 bis 12/2004) widerspiegelt, in denen unterschiedliche Beteiligungsformen wissenschaftlicher Begleitung und somit verschiedene Formen des Aufgabenverständnisses zum Ausdruck kommen. Hier gilt "[...], dass diese Bandbreite bei den Aufgabendefinitionen konzeptionell toleriert werden oder sogar erwünscht sein kann, aber [...] es gälte zu prüfen, ob ein solches Konzept eine taugliche Basis für eine effektive Entwicklungsarbeit darstellt." (Ebner, Pätzold 2003: 221).

Hierzu ist jedoch zunächst ein Einstellungswechsel unabdingbar, der seinen Ausdruck in der Erkenntnis findet, dass wissenschaftliche Begleitung, vor allem aber Evaluation eben nicht 'Zutat' (vgl. Ploghaus 1995: 106), sondern Voraussetzung für den Transfer der Modellversuchsergebnisse ist. Die BLK bzw. das BMBF als übergeordnete Instanzen verstehen die wissenschaftliche Begleitung als einen Aspekt, auf den höchsten Wert gelegt wird; diese Wertschätzung sollte sich auf Grund der Ergebnisse einzelner wissenschaftlicher Begleitungen, die sich lediglich in der abschließenden Dokumentation nicht nachvollziehbarer 'Evaluations-Resultate' niederschlagen und der stark kritikbehafteten Tätigkeiten der Mehrzahl der untersuchten Ergebnisse wissenschaftlicher Begleitungen durch stärkere Präsenz und anschließende Ergebnisprüfung ausdrücken. Vor dem Hintergrund des personellen und finanziellen Aufwandes, der mit den Erprobungsvorhaben verbunden ist, scheint diese Forderung gerechtfertigt.

Im Hinblick auf die modellversuchs- bzw. evaluationsbezogene Ebene sollten folgende Vorschläge diskutiert werden, die sich an den Evaluationsstandards des Joint Committee for Educational Evaluation orientieren. Damit soll die Notwendigkeit der an Qualitätsstandards orientierten Weiterentwicklung der Evaluationsforschung und die Anwendung der Erkenntnisse hieraus hervorgehoben werden. Ergebnis sollte eine auf elaborierter Evaluationsforschung basierende (theoriegeleitete) Evaluationspraxis sein, um so die Qualität der transferierbaren Ergebnisse steigern zu können:

Vollständige Transparenz aller MV-Tätigkeiten, hergestellt durch die funktionale Gestaltung der Modellversuchsdokumentationen, d. h. unter anderem ...

- verfassen unterschiedlicher, d. h. adressatendifferenzierter Berichtsarten.
- Einsatz solcher Evaluationsinstrumente, die nur minimale Störungen im Modellversuchsfeld mit sich bringen.
- Offenlegung aller Interessenkonflikte und Probleme bei der Durchführung des Modellversuchs.
- Installation eines Informationsservers, der einen uneingeschränkten Zugriff auf die erstellten Dokumentationen gewährleistet.
- Herstellen einer einheitlich hohen Qualität wissenschaftlicher Begleitung/Evaluation, d. h. Orientierung an anerkannten Evaluationsstandards.
- Durchführung von Metaevaluationen durch externe Mitarbeiter des Programmträgers.

Um die Informationsbedürfnisse der vorgesehenen Evaluationsnutzer zu befriedigen (Nützlichkeits-Standards), ist neben der Ermittlung der Beteiligten und Betroffenen auf vollständige Transparenz durch die entsprechende Gestaltung der Modellversuchsdokumentationen zu achten. Diese Transparenz drückt sich u. a. in der Nachvollziehbarkeit des Vorgehens bei der Gestaltung des Modellversuchsfeldes, d. h. Aufgabenverteilung, Entwicklung der Modellversuchsunterlagen, Offenlegung der Auswirkungen des Modellversuchs auf die Beteiligten und Betroffenen und der Evaluation aus.

Auf Grund der unterschiedlichen Nutzergruppen (Schulpraktiker und zukünftige wiss. Begleitungen) der Modellversuchsdokumentationen müssen verschiedene Berichtsarten verfasst werden, die sich hinsichtlich Sprache und inhaltlichen Schwerpunkten unterscheiden. Auch wenn hiermit eine Steigerung des Arbeitsaufwands einhergeht, so ist dieses Vorgehen unumgänglich, wenn die Akzeptanz der Modellversuchsergebnisse gesteigert und sichergestellt werden soll, dass aus einem Explikationsmodellversuch auch ein Planungsmodellversuch wird, d. h. der Transfer der Ergebnisse vorgenommen wird.

Des Weiteren haben die mit der Evaluation beauftragten Mitarbeiter der wissenschaftlichen Begleitung bei der Auswahl und dem Einsatz ihrer Evaluationsinstrumente sicherzustellen (Durchführbarkeitsstandards), dass diese derart gestaltet

sind, dass erhebliche Störungen auf der Objektebene vermieden werden. Sollte diese Anforderung verletzt werden, entsteht der Eindruck der permanenten Kontrolle bei den Mitarbeitern im Modellversuchsfeld, so dass Ablehnungen oder angepasstes Verhalten die Folge sein können.

Im Rahmen der Korrektheitsstandards sind es zwei Aspekte, die von besonderer Bedeutung sind. Zum einen sind Interessenkonflikte oder Probleme bei der Entwicklung und Ausgestaltung des Modellversuchsfeldes offen zu legen, damit diese nach dem Transfer vermieden werden können. Vor allem aber ist den Informationsbedürfnissen des ganzen Zielgruppenspektrums Rechnung zu tragen. D. h. eine Veröffentlichungspflicht, die mit der Übergabe oder Verbreitung durch den Programmträger erfüllt ist, reicht nicht aus. Vielmehr sollte ein zentraler Informationsserver bei der BLK installiert werden, auf dem alle Modellversuchsberichte abgelegt werden, um einen Zugriff durch Interessierte jederzeit zu ermöglichen.

Wesentliches Ziel möglicher Veränderungen auf Ebene der Evaluation ist jedoch das Sicherstellen einer einheitlich hohen Qualität der Funktionen wissenschaftlicher Begleitung (Beratung resp. formative Evaluation sowie summativ Evaluation der evaluationsbezogenen Zielsetzungen). Hierfür bieten schwerpunktmäßig die Genauigkeitsstandards Orientierungshilfe, da sie sicherstellen, dass 'fachlich angemessene' Informationen über die Güte und Verwendbarkeit des Modellversuchs generiert werden (vgl. Sanders 1999: 155). Eine weitere Ausgestaltung der Evaluationsforschung, deren Ergebnisse dann auch von versierten Evaluatoren in die Evaluationspraxis getragen werden, ist hierfür unabdingbar.

Literaturverzeichnis

- Achtenhagen, F. (1973). Methodologische Probleme empirischer Begleituntersuchungen zu pädagogischen Innovationsversuchen unter statistischem Aspekt. *Zeitschrift für Pädagogik* 13, H. 1, S. 43-62.
- Arnoff, S. (1987). Evaluation issues in the educational product live cycle. In L. S. May (Hrsg.), *Evaluating business and industry training*. S. 94-134, Boston, Dordrecht, Lancaster.
- Bader, R. (1992). Verbindung von Aus- und Weiterbildung – Aspekte für Modellversuche in der beruflichen Bildung. In P. Dehnhostel (Hrsg.), *Lernen für die Zukunft durch verstärktes Lernen am Arbeitsplatz*, S. 225-241, Berlin.
- Bähr, W. & Holz, H. (1995). *Was leisten Modellversuche*. Bonn.
- Bea, F.-X., Dichtl, E. & Schweitzer, M. (1991). *Allgemeine Betriebswirtschaftslehre*. Stuttgart.
- Beck, U.; Bonß, W. (1991). Verwendungsforschung – Umsetzung wissenschaftlichen Wissens. In U. Flick (Hrsg.), *Handbuch qualitative Sozialforschung: Grundlagen, Konzepte, Methoden und Anwendung*, S. 416-419, München.
- Benteler, P. (1991). Möglichkeiten und Grenzen der Übertragung von Modellversuchsergebnissen. *Berufsbildung in Wissenschaft und Praxis*, 10, H. 6, S. 7-12.
- Beywl, W. (1988). *Zur Neufassung der Evaluationsmethodologie*. Köln.
- Beywl, W. (1991). Entwicklung und Perspektiven praxiszentrierter Evaluation. *Sozialwissenschaften und Berufspraxis*, H3, S. 265-279.
- Bickman, L. (1987). *Using Program Theory in Evaluation. New Directions for Program Evaluation*. San Francisco.
- Bortz, J. (1995). *Forschungsmethoden und Evaluation*. Berlin.
- Bremen, Freie Hansestadt – Senator für Bildung und Wissenschaft (1994, 1995, 1996). *Modellversuch Umweltbildung in Berufen der öffentlichen Verwaltung. Entwicklung und Erprobung ökologischer Lernangebote für die berufliche Aus- und Weiterbildung*. Erster und zweiter Zwischenbericht, Abschlussbericht. Bremen.
- Bremen, Freie Hansestadt – Senator für Bildung und Wissenschaft (1999, 2001). *Modellversuch Kundenorientierung in der kaufmännischen Berufsausbildung –KUNO–*, Zwischen- und Abschlussbericht. Bremen.
- Brezinka, W. (1971). Über Erziehungsbegriffe. *Zeitschrift für Pädagogik*. 17, H. 5, S. 567-615.
- Bundesinstitut für Berufsbildung (1978). *Modellversuche - ein Instrument zur Weiterentwicklung beruflicher Bildungspraxis*. Berlin.
- Bundesinstitut für Berufsbildung (1985). *Modellversuche im außerschulischen Bereich der Berufsbildung*. Unveröffentlichtes Typoskript. Berlin.
- Bundesinstitut für Berufsbildung (1997). *Modellversuche in der außerschulischen Berufsbildung*. Berlin.

- Bundesministerium für Bildung und Wissenschaft (1970). *Bildungsbericht 1970. Bericht der Bundesregierung zur Bildungspolitik*. Bonn.
- Bundesministerium für Bildung und Wissenschaft (1982). *Informationen Bildung und Wissenschaft*. H. 11, S. 197 – 200.
- Bundesministerium für Bildung und Wissenschaft (1984). *Informationen Bildung und Wissenschaft*. H. 9, S. 149 – 155.
- Bundesministerium für Bildung und Wissenschaft (1985). *Informationen Bildung und Wissenschaft*. H. 2, S. 25 – 27.
- Bundesministerium für Bildung und Wissenschaft (1985). *Informationen Bildung und Wissenschaft*. H. 3, S. 44 – 46.
- Bundesministerium für Bildung und Wissenschaft (1985). *Informationen Bildung und Wissenschaft*. H. 7-8, S. 123 – 125.
- Bundesministerium für Bildung und Wissenschaft (1986). *Informationen Bildung und Wissenschaft*. H. 7-8, S. 99 – 103.
- Bundesministerium für Bildung und Wissenschaft (1987). *Informationen Bildung und Wissenschaft*. H. 3, S. 41 – 45.
- Bundesministerium für Bildung und Wissenschaft (1987). *Informationen Bildung und Wissenschaft*. H. 5, S. 75 – 77.
- Bundesministerium für Bildung und Wissenschaft (1987). *Informationen Bildung und Wissenschaft*. H. 6, S. 91 – 101.
- Bundesministerium für Bildung und Wissenschaft (1987). *Informationen Bildung und Wissenschaft*. H. 10, S. 148 – 149.
- Bundesministerium für Bildung und Wissenschaft (1987). *Informationen Bildung und Wissenschaft*. H. 12, S. 181 – 182.
- Bundesministerium für Bildung und Wissenschaft (1988). *Informationen Bildung und Wissenschaft*. H. 3, S. 35 – 41.
- Bundesministerium für Bildung und Wissenschaft (1988). *Informationen Bildung und Wissenschaft*. H. 11, S. 47 – 49.
- Bundesministerium für Bildung und Wissenschaft (1989). *Informationen Bildung und Wissenschaft*. H. 3, S. 31 – 35.
- Bundesministerium für Bildung und Wissenschaft (1990). *Informationen Bildung und Wissenschaft*. H. 2, S. 9 – 11.
- Bundesministerium für Bildung und Wissenschaft (1991). *Informationen Bildung und Wissenschaft*. H. 1, S. 3 – 5.
- Bundesministerium für Bildung und Wissenschaft (1991). *Informationen Bildung und Wissenschaft*. H. 2, S. 9 – 10.
- Bundesministerium für Bildung und Wissenschaft (1991). *Informationen Bildung und Wissenschaft*. H. 3, S. 26 – 29.

- Bundesministerium für Bildung und Wissenschaft (1991). *Informationen Bildung und Wissenschaft*. H. 4.
- Bundesministerium für Bildung und Wissenschaft (1991). *Informationen Bildung und Wissenschaft*. H. 5.
- Bundesministerium für Bildung und Wissenschaft (1992). *Informationen Bildung und Wissenschaft*. H. 2, S. 9 – 59.
- Bundesministerium für Bildung und Wissenschaft (1992). *Informationen Bildung und Wissenschaft*. H. 12, S. 162 – 163.
- Bundesministerium für Bildung und Wissenschaft (1993). *Informationen Bildung und Wissenschaft*. H. 2, S. 9 – 11.
- Bundesministerium für Bildung und Wissenschaft (1993). *Informationen Bildung und Wissenschaft*. H. 7-8, S. 91.
- Bundesministerium für Bildung und Wissenschaft (1993). *Informationen Bildung und Wissenschaft*. H. 11, S. 146 – 147.
- Bundesministerium für Bildung und Wissenschaft (1993). *Informationen Bildung und Wissenschaft*. H. 12, S. 164.
- Bundesministerium für Bildung und Wissenschaft (1994). *Informationen Bildung und Wissenschaft*. H. 1-2, S. 2.
- Bundesministerium für Bildung und Wissenschaft (1994). *Informationen Bildung und Wissenschaft*. H. 3, S. 17 – 21.
- Bundesministerium für Bildung und Wissenschaft (1994). *Informationen Bildung und Wissenschaft*. H. 6, S. 59.
- Bundesministerium für Bildung und Wissenschaft (1994). *Informationen Bildung und Wissenschaft*. H. 7-8, S. 75.
- Bundesministerium für Bildung und Wissenschaft (1994). *Informationen Bildung und Wissenschaft*. H. 9, S. 97 – 101.
- Bundesministerium für Bildung und Wissenschaft (1994). *Informationen Bildung und Wissenschaft*. H. 10, S. 114 – 119.
- Bundesministerium für Bildung und Wissenschaft (1994). *Informationen Bildung und Wissenschaft*. H. 11, S. 130 – 132.
- Bundesministerium für Bildung und Wissenschaft (1997). *Prospect*. H. 1 und 2 und 3.
- Bundesministerium für Bildung und Wissenschaft (1998). *Prospect*. H. 1 und 2 und 3.
- Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (1994). *Berufsbildungsförderungsgesetz*.
- Bundesministerium für Bildung und Forschung (2001a). *Innovations- und Transfereffekte von Modellversuchen in der beruflichen Bildung*. Bonn.
- Bundesministerium für Bildung und Forschung (2001b). *Berufsbildung für eine nachhaltige Entwicklung*. Bonn.

- Bundesregierung: *Bulletin* (1983), H. 15; (1983), H. 17; (1983), H. 121; (1984), H. 152; (1984), H. 85; (1985), H. 45; (1985), H. 84; (1985), H. 88; (1986), H. 100; (1986), H. 135; (1987), H. 33; (1988), H. 39; (1988), H. 122; (1989), H. 26; (1989), H. 29; (1989), H. 44; (1991), H. 42; (1992), H. 118; (1994), H. 35; (1994), H. 119.
- Bundeszentrale für politische Bildung (1992). *Bildungspolitik in Deutschland 1945-1990. Ein historisch-vergleichender Quellenband*. Bonn.
- Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (1976). *Informationen über die Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung*. Bonn.
- Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (1981). *10 Jahre Modellversuche im Bildungswesen*. Bonn.
- Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (1995). *Modellversuche in der Bewährung. 2. Bericht zur Umsetzung von Modellversuchen im Bildungswesen*. Bonn.
- Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (1997). *Informationsschrift über gemeinsame Modellversuchsförderung im Bildungswesen durch Bund und Länder*. Bonn.
- Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (1998). *Informationen über Modellversuchsförderung durch Bund und Länder im Rahmen der BLK ab 01.01.1998*. <http://www.k.shuttle.de/blk/mv-innovativ.htm>.
- Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (1986). *Ausgaben der Gebietskörperschaften für Bildung und Wissenschaft in den Jahren 1970 bis 1985*. Bonn.
- Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (1996). *Ausgaben der Gebietskörperschaften für Bildung und Wissenschaft in den Jahren 1993 (Ist), 1994 und 1995 (Soll)*. Bonn.
- Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung. *Jahresberichte 1988, 1990, 1991, 1993, 1994, 1995, 1996, 1997*.
- Chen, H. – T. & Rossi, P. (1980). The Multi-Goal, Theory Driven Approach to Evaluation – A Model Linking Basic and Applied Social Science. *Social Forces*, 59, S. 106-122.
- Chen, H. –T. (1990). *Theory-Driven Evaluations*. Newsbury Park.
- Christlich Demokratische Union Deutschlands: *Die Wahlprogramme der CDU und CSU 1976, 1980, 1983, 1987, 1990, 1994*. Bonn.
- Clarke, A. (1999). *Evaluation Research - An Introduction to Principles, Methods and Practice*. London.
- Cronbach, L. J. (1972). Evaluation zur Verbesserung von Curricula. In C. Wulf, *Evaluation*. S. 41 – 60, München.
- Cronbach, L. J. (1973). *Course improvement through evaluation. School evaluation – The politics and process*. Berkley.
- Cronbach, L. J. (1980). *Toward Reform of Program Evaluation*. San Francisco.

- Cronbach, L. J. & Snow, R. (1981). *Aptitudes and instructional methods : a handbook for research and interactions*. New York.
- Cronbach, L. J. (1982). *Designing Evaluations of Educational and Social Programs*. San Francisco.
- Dannhäuser, A. u. a. (1988). *Ist die Schule noch zu retten? Plädoyer für eine neue Bildungsreform*. Weinheim, Basel.
- Dehnbostel, P. u. a. (1992). *Neue Technologien und berufliche Bildung. Berichte zur beruflichen Bildung*. Hrsg. BiBB. Berlin.
- Dehnbostel, P. u. a. (1994). Modellversuche als Innovationen in der Berufsbildung. *Berufsbildung*, 48, H. 30, S. 3-7.
- Dehnbostel, P.: Neuorientierungen wissenschaftlicher Begleitforschung – eine kritische Auseinandersetzung mit bestehenden Konzepten und fälligen Fortentwicklungen. In P. Benteler (Hrsg.), *Modellversuchsforschung als Berufsbildungsforschung*, S. 71-98, Köln.
- Dehnbostel, P. & Demuth, B. (1995). Wissenschaftliche Begleitforschung und Transfer in der Modellversuchsreihe 'Dezentrales Lernen'. In BiBB (Hrsg.), *Was leisten Modellversuche?* S. 43-52, Berlin.
- Dehnbostel, P. (1998). Begleitforschung von Modellversuchen zwischen Praxisinnovation und Theorieentwicklung. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, H. 2, S. 187-203.
- Deutscher Bundestag (1967). *Bericht über den Stand der Nachbesserungen auf dem Gebiet der Bildungsplanung*. Drucksache V/2166.
- Dewey, J. (1929). *The Quest for Certainty*. New York.
- Dewey, J. (1933). *How We Think*. New York.
- Diener, W. & Zimmer, G. (1995). Schlussfolgerungen für die Modellversuchsreihe 'Multimediales Lernen in der Berufsbildung'. In H. Holz (Hrsg.), *Was leisten Modellversuche?* S. 277-294, Berlin.
- Ebner, H. – G. (1989). Über den Umgang mit Theorien in Zusammenhang mit der theoretischen Fundierung handlungsorientierten Lernens in der Didaktik der Wirtschaftslehre. In R. Czycholl, H. – G. Ebner (Hrsg.), *Zur Kritik handlungsorientierter Ansätze in der Didaktik der Wirtschaftslehre*. S. 115-164, Berlin.
- Ebner, H. - G. (1992). Facetten und Elemente didaktischer Handlungsorientierung. In G. Pätzold (Hrsg.), *Handlungsorientierung in der beruflichen Bildung*, S. 33-55, Frankfurt am Main.
- Ebner, H. - G. & Pätzold, G. (2003). Berufsbildungsforschung im BLK-Modellversuchsprogramm "Innovative Konzepte der Lehrerbildung (2. und 3. Phase) für berufsbildende Schulen" (innovelle-bs). *Zeitschrift für Berufs- und Wirtschaftspädagogik*, H.2, S: 213 – 221.
- Ehrlich, K. (1995). Auf dem Weg zu einem neuen Konzept wissenschaftlicher Begleitung. *Berufsbildung in Wissenschaft und Praxis*. 24, H.1. S. 32-37.

- Eickhorst, A. (1977). Der Innovationsbegriff – Ergebnisse einer Bestandsaufnahme und Inhaltsanalyse. *Unterrichtswissenschaft*, 5, H. 2, S. 174-183.
- Eisner, E. W. (1983). Educational Connoisseurship and Criticism. Their Form und Function in Educational Evaluation. In G. F. Madaus (Hrsg.), *Evaluation Models. Viewpoints on Educational and Human Services Evaluation*, S. 335-347, Boston.
- Ellwein, T. (1964). Die verwaltete Schule. *Das Argument*, 6, H. 4, S. 209-220.
- Euler, D. (1994). *Didaktik einer sozio-informationstechnischen Bildung*. Köln.
- Euler, D. (1995a): Organisationsentwicklung – Eine neue Perspektive für die Modellversuchsarbeit in der beruflichen Bildung: In G. Dybowski (Hrsg.), *Berufsbildung und Organisationsentwicklung*, S. 240-257, Bremen.
- Euler, D. (1995b). Transfer von Modellversuchsergebnissen: Theoretische Fundierung, empirische Hinweise und erste Konsequenzen. In P. Benteler (Hrsg.), *Modellversuchsforschung als Berufsbildungsforschung*, S. 225-268, Köln.
- Euler, D. (1996). Denn sie tun nicht, was sie wissen. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, H. 4, S. 350-365.
- Euler, D. & Dehnborstel, P. (1998). Berufliches Lernen als Forschungsgegenstand – Inwieweit sind Modellversuche in der Berufsbildung ein Instrument von Erkenntnisgewinnung und Praxisinnovation? In D. Euler (Hrsg.), *Berufliches Lernen im Wandel – Konsequenzen für die Lernorte?* S. 490-500. Nürnberg.
- Euler, D. & Kröll, Twardy, M. (1988): Modellversuch „Pädagogische Beratung im Handwerk“. *Berufsbildung in Wissenschaft und Praxis*, 17, H. 1, S. 29-34.
- Euler, D. & Kutt, K. (1995). Transfer von Modellversuchsergebnissen: Bedingungen und Hinweise für die Gestaltung des Transferprozesses. In P. Benteler (Hrsg.), *Modellversuchsforschung als Berufsbildungsforschung*, S. 269-294, Köln.
- Euler, D. & Sloane, P. F. E. (1993). Modellversuchsforschung. Überlegungen zu einem wirtschaftspädagogischen Forschungsansatz. *Erziehungswissenschaft und Beruf – Vierteljahresschrift für Unterrichtspraxis und Lehrerbildung*, 41, H. 2, S. 203-205.
- Euler, D. & Sloane, P. F. E. (1998). *Implementation als Problem der Modellversuchsforschung*. <<http://www.wiso.uni-erlangen.de/WiSo/BWi/WiPd/seiten/virtpubl.html>>.
- Fichtner, D. (1973). Bildungsplanung und Bildungspolitik des Bundes. *Aus Politik und Zeitgeschichte*, H. 39, S. 17-29.
- Flitner, A. (1977). *Universität heute: Wem dient sie? Wer steuert sie?* München.
- Flodan, R. E. (1983). Accreditation and Evaluation. In G. F. Madaus (Hrsg.), *Evaluation Models. Viewpoints on Educational and Human Services Evaluation*, S. 261-277, Boston.
- Freiheitlich Demokratische Partei Deutschlands (1977). *Programmatistische Entwicklung der FDP 1946 bis 1969. Einführung und Dokumente*. Meisenheim am Glan.
- Freiheitlich Demokratische Partei Deutschlands (1980). *Wahlprogramm 1980*. Freiburg.
- Frey, K. (1972). *Curriculum-Handbuch*. München.

- Friedeburg, L. v. (1992). *Bildungsreform in Deutschland. Geschichte und gesellschaftlicher Widerspruch*. Frankfurt am Main.
- Führ, C. (1967). *Schulversuche 1965/66. Teil I: Gesamtdarstellung*. Frankfurt am Main.
- Führ C. (1997). *Deutsches Bildungswesen seit 1945. Grundzüge und Probleme*. Neuwied, Kriftel, Berlin.
- Fullan, M. (1993). Coordinating school and district development in restructuring. In J. Murphy & P. Hallinger (Hrsg.), *Restructuring Schools – Learning from Ongoing Efforts*. Newbury Park.
- Fullan, M. (1999). *Die Schule als lernendes Unternehmen. Konzepte in einer neuen Kultur der Pädagogik*. Stuttgart.
- Deutsche Forschungsgemeinschaft (1990). *Berufsbildungsforschung an den Hochschulen der Bundesrepublik Deutschland*. Weinheim.
- Glass, G. V. (1977). Die Entwicklung der Methodologie der Evaluation. In C. Wulff, *Theorie und Konzepte der Erziehungswissenschaft*. München.
- Goldstein, I. R. (1987). Content validity as an evaluation strategy for examining training programs. In L. S. May (Hrsg.), *Evaluating business and industry training*, S. 94-134, Boston.
- Gronlund, N. E. & Linn, R. L. (1990). *Measurement and evaluation in teaching*. New York.
- Grüner, H. (1993). Evaluation und Evaluationsforschung im Bildungswesen. *Pädagogische Rundschau*, H. 1, S. 29-52.
- Grüner, H. (1995). Zur Entwicklung der Evaluationsforschung im Bildungswesen. In K. – H. Sommer (Hrsg.), *Problemfelder der Berufs- und Wirtschaftspädagogik*, S. 382-400, Esslingen.
- Guba, E. & Lincoln, Y. S. (1989). *Fourth generation evaluation*, London.
- Haan G. de & Hamm-Brücher, H., Reichel, N. (2000). *Bildung ohne Systemzwänge – Innovationen und Reformen*. Neuwied.
- Habermas, J. (1969). *Protestbewegung und Hochschulreform*. Frankfurt.
- Habermas, J. (1974). *Theorie der Gesellschaft oder Sozialtechnologie – Was leistet die Systemforschung?* Frankfurt am Main.
- Hamburg, Freie und Hansestadt – Amt für Schule (1992, 1993, 1994). *Schlüsselqualifikationen für Kaufleute im Einzelhandel*. Erster und zweiter Zwischenbericht, Abschlussbericht, Hamburg.
- Hamm-Brücher, H. (1972). *Unfähig zur Reform? Kritik und Initiativen zur Bildungspolitik*. München.
- Heimerer, L. (1998). *Evaluation der schulischen Ausbildung "Kaufmann/Kauffrau im Einzelhandel". Ergebnisse der Untersuchungen an 19 Schulen in Bayern, Berlin, Nordrhein-Westfalen und Sachsen*. Bielefeld.
- Heinze, T. (1981). *Praxisforschung. Wie Alltagshandeln und Reflexion zusammengebracht werden können*. München.

- Helberger, C. & Palamidis, H. (1989). Der Beitrag der Humankapitaltheorie zur Erklärung der Bildungsnachfrage. In P. Döring (Hrsg.), *Bildung in sozioökonomischer Sicht*, S. 205-220, Köln.
- Henry, G. T. (2001). Coeditor's Notes. In D. L. Stufflebeam (Hrsg.), *Evaluation Models*, San Francisco.
- Hessen, Bundesland (1998, 1999, 2000). *Selbstorganisierte Lernprozesse und neue Lernwelten in der beruflichen Bildung*. Erster und zweiter Sachbericht, Abschlussbericht.
- Hoepke, I. & Wolf, B. (1992). Ansätze zur Untersuchung von Antezedenzbedingungen für die Übertragung von Modellversuchsergebnissen. In P. Dehnbostel (Hrsg.), *Neue Technologien und berufliche Bildung*, Berlin.
- House, E. R. (1980). *Evaluation with Validity*. Beverly Hills.
- House, E. R. (1983). Assumptions Underlying Evaluation Modells. In G. F. Madaus (Hrsg.), *Evaluation Models. Viewpoints on Educational and Human Services Evaluation*, S. 45-64, Boston.
- Houston, T. R. (1972). The Behavioural Sciences. Impact-Effectiveness Model. In P. Rossi (Hrsg.), *Evaluating Social Programs. Theory, Practice and Politics*. New York.
- Huisken, F. (1972). *Zur Kritik bürgerlicher Didaktik und Bildungsökonomie*. München.
- Huschke-Rhein, R. (1987). *Systempädagogische Wissenschafts- und Methodenlehre. Qualitative Forschungsmethoden und Handlungsforschung*. Köln.
- Joint Committee on Standards for Educational Evaluation (1999). *Handbuch der Evaluationsstandards*. J. R. Sanders (Hrsg.), Opladen.
- Jongebloed, H. – C. & Twardy, M. (1983). Wissenschaftstheoretische Voraussetzungen. In M. Twardy (Hrsg.), *Konsumpädagogik*, S. 1-73, München, Köln.
- Kaiser, F. – J. & Weitz, B. O. (1990). *Arbeiten und Lernen in schulischen Modellunternehmen. Neue Informationstechnologien und Datenverarbeitung im Berufsfeld Wirtschaft und Verwaltung. Verbindung von berufspraktischer und theoretischer Arbeit (Ausbildung) im Lernbüro*. Bad Heilbrunn/Obb.
- Kieser, A. (2001). *Organisationstheorien*. Köln.
- Klafki, W. (1973). Handlungsforschung im Schulfeld. *Zeitschrift für Pädagogik*, 19, H. 4, S. 487-516.
- Klöti, U. & Widmer, T. (1997). Untersuchungsdesigns. In W. Bussmann (Hrsg.), *Einführung in die Politikevaluation*, S. 185-213, Basel.
- Knauss, G. (1978). Modellversuche im Bildungswesen als Instrument einer gemeinsamen Bildungsplanung von Bund und Ländern. In BiBB (Hrsg.), *Modellversuche – Ein Instrument zur Weiterentwicklung beruflicher Bildungspraxis*, S. 221-227, Berlin.
- Koch, P. u. a. (1979). *Determinanten staatlicher Bildungsplanung und -politik 1969-1978*. Köln.
- Kordes, H. (1995). Pädagogische Aktionsforschung. In H. Haft (Hrsg.), *Enzyklopädie Erziehungswissenschaft*, S. 185-219, Stuttgart.

- Kreuser, K. (1988). Bildungsreform und Föderalismus. In A. Dannhäuser (Hrsg.), *Ist die Schule noch zu retten?*, S. 135-153, Weinheim und Basel.
- Kromrey, H. (2000). *Empirische Sozialforschung*. Opladen.
- Kromrey, H. (2001). Evaluation – ein vielschichtiges Konzept. *Sozialwissenschaft und Berufspraxis*, 24, S. 105-131.
- Kuhn, T. (1969, 1973, 1981). *Die Struktur wissenschaftlicher Revolutionen*. Frankfurt am Main.
- Kutt, K. (1995). Modellversuche und Organisationsentwicklung. In G. Dybowski (Hrsg.), *Berufsbildung und Organisationsentwicklung*, S. 240-257, Bremen.
- Lakatos, I. (1974). The Methodology of Scientific Research Programmes: In J. Worrall (Hrsg.), *Philosophical Papers, Volume 1*, Cambridge.
- Lamnek, S. (1995a). *Qualitative Sozialforschung. Methodologie*. Weinheim.
- Lamnek, S. (1995b). *Qualitative Sozialforschung. Methoden und Techniken*. Weinheim.
- Laur, U. (1978). Forschungsstrategien in Modellversuchen – Das Beispiel MME. In BiBB (Hrsg.), *Modellversuche – Ein Instrument zur Weiterentwicklung beruflicher Bildungspraxis*, S. 30-41, Berlin.
- Lewin, K. (1963). *Feldtheorie in den Sozialwissenschaften*. Bern.
- Lipsey, M. (1987). *Theory as Method: Small Theories of Treatments. Paper presented at the National Center for Health Services Research Conference, "Strengthening Causal Interpretations of Non-Experimental Data"*. Tucson.
- Lipsmeier, A. (1997). Zur wissenschaftlichen Begleitung von CAL-Modellversuchen im Allgemeinen und zur Evaluation von 'OLLI' im Besonderen. *Berufsbildung in Wissenschaft und Praxis*, 26, H. 1, S. 22-27.
- Menck, P. (1978). Innovation. In C. Wulf (Hrsg.), *Wörterbuch der Erziehung*. S. 305-308, München.
- Mertens, D. (2000). *Transformative theory and evaluation. Theories of Evaluation TIG Newsletter. Theories of Evaluation Topical Interest Group of the American Evaluation Association*. New York.
- Meyer-Dohm, P. (1991). *Theoretische Grundlagen einer neuen Berufsbildung*. Stuttgart.
- Miles, M. (1971). *Innovation in Education*. New York.
- Mitter, W. & Weishaupt, H. (1977). *Ansätze zur Analyse der wissenschaftlichen Begleitung bildungspolitischer Innovationen*. Frankfurt am Main.
- Moser, H. (1995). *Grundlagen der Praxisforschung*. Freiburg im Breisgau.
- Nickolaus, R. & Schurpel, U. (2001a). *Berufsbildung für eine nachhaltige Entwicklung. Bundesministerium für Bildung und Forschung* (Hrsg.), Berlin.
- Nickolaus, R. & Schurpel, U. (2001b). *Innovations- und Transfereffekte von Modellversuchen in der beruflichen Bildung. Bundesministerium für Bildung und Forschung* (Hrsg.), Berlin.

- Nordrhein-Westfalen, Bundesland (1996/1997, 1997/1998, 1999). *Modellversuch EUWAS – Integration einer europäischen Dimension in die kaufmännische Berufsausbildung. Der BLK-Modellversuch „Euro-Wirtschaftsassistent/-in“ (EUWAS)*. Erster und zweiter Zwischenbericht, Abschlussbericht.
- Novak, Hermann (1995). Transfer von Modellversuchsergebnissen – Von der "Restgröße" am Ende der Laufzeit zum laufzeitsimultanen Lern- und Implementierungsprozess. In W. Bähr (Hrsg.), *Was leisten Modellversuche?*, S. 53-66, Berlin.
- OECD (1996). *Wissensgrundlagen für die Bildungspolitik*. Frankfurt am Main.
- OECD (1998). *Die Internationalisierung des Hochschulwesens – ein OECD/CERI-Bericht*. Frankfurt am Main.
- Opp, K. – D. (1970, 1976). *Methodologie der Sozialwissenschaften – Einführung in Probleme ihrer Theoriebildung*. Hamburg.
- Owens, T. (1973). Educational Evaluation by Adversary Proceedings. In E. R. House (Hrsg.), *School evaluation – The politics and process*, Berkley.
- Patton, M. Q. (1997). *Utilization-focused evaluation – The new century text*. London.
- Pätzold, G. (1995). Ansprüche an die pädagogische Begleitforschung im Rahmen von Modellversuchen. In P. Benteler (Hrsg.), *Modellversuchsforschung als Berufsbildungsforschung*, S. 45-70, Köln.
- Picht, G. (1964). *Die deutsche Bildungskatastrophe: Analyse und Dokumentation*. München.
- Ploghaus, G. (1992). Zwanzig Jahre Modellversuche in der Berufsbildung. In H. Pütz (Hrsg.), *Innovationen in der beruflichen Bildung*, S. 209-218, Berlin und Bonn.
- Ploghaus, G. (1995). Bedeutung der Modellversuchs-Evaluation und Erwartungen aus der Sicht des BMBF. In P. Benteler (Hrsg.), *Modellversuchsforschung als Berufsbildungsforschung*, S. 101-112, Köln.
- Popham, W. J. (1988). *Educational evaluation*. Englewood.
- Posavac, E. J. & Carey, R. G. (1989). *Program evaluation*. Englewood.
- Prim, R. & Tilmann, H. (1979). *Grundlagen einer kritisch-rationalen Sozialwissenschaft - Studienbuch zur Wissenschaftstheorie*. Heidelberg.
- Raschert, J. (1980). Bildungspolitik im kooperativen Föderalismus. Die Entwicklung der länderübergreifenden Planung und Koordination des Bildungswesens der Bundesrepublik Deutschland. In Max-Planck-Institut für Bildungsforschung (Hrsg.), *Bildung in der Bundesrepublik Deutschland*, S. 103-216, Berlin.
- Recum, H. v. (1972). Das Bildungswesen und der Entwicklungsprozess. In A. O. Schorb (Hrsg.), *Bildungsplanung und Bildungspolitik*, S. 45-58, Frankfurt am Main.
- Rolff, H.-G. (1971). *Bildungsplanung als rollende Reform: eine soziologische Analyse der Zwecke, Mittel und Durchführungsnormen einer reformbezogenen Planung des Bildungswesens*. Frankfurt am Main.
- Rossi, P. H. & Freeman, H. E. (1982, 1986, 1995). *Evaluation. A systematic approach*. Beverly Hills, London, New Dehli.

- Roth, H.-G. (1975). *25 Jahre Bildungsreform in der Bundesrepublik. Bilanz und Perspektiven*. Bad Heilbronn.
- Rottluff, J. (1992). Möglichkeiten und Grenzen der Erfassung des Transfers von Modellversuchsergebnissen – Fragen und Thesen. In P. Dehnbostel (Hrsg.), *Neue Technologien und berufliche Bildung*, S. 265-269, Berlin.
- Rügemer, W. (1991). Akademische Verdummung oder zivile Intellektualität. In W. Rügemer (Hrsg.), *In der deutschen Bildungs-Spirale*, S. 5-14, Köln.
- Saarland, Bundesland (1995). *Entwicklung und Erprobung ganzheitlicher Lernansätze in der kaufmännischen Berufsausbildung am Beispiel des Ausbildungsberufs Versicherungskaufmann/-frau*. Zwischenbericht, Abschlussbericht.
- Schaub, H. & Zenke, K. (1995). *Wörterbuch zur Pädagogik*. München.
- Schmiel, M. & Sommer, K.-H. (1985). *Lehrbuch Berufs- und Wirtschaftspädagogik*. München.
- Schnell, R., Hill, P. & Esser, E. (1999). *Methoden der Sozialforschung*. München.
- Scriven, M. (1971). Objectivity and Subjectivity in Educational Research. In L. G. Thomas (Hrsg.), *Philosophical Redirection of Educational Research*, Chicago.
- Scriven, M. (1972). Die Methodologie der Evaluation. In C. Wulf (Hrsg.), *Evaluation*, S. 60-91, München.
- Scriven, M. (1973). Goal-free evaluation. In E. R. House (Hrsg.), *School evaluation – The politics and process*, S. 319-328, Berkley.
- Scriven, M. (1991). *Evaluation Thesaurus*. Newbury Park.
- Sekretariat der Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (1977). *Handbuch für die Kultusministerkonferenz*. Bonn.
- Sloane, P. F. E. (1992). *Modellversuchsforschung*. Köln.
- Sozialdemokratische Partei Deutschlands (1959). *Grundsatzprogramm der Sozialdemokratischen Partei Deutschlands*. Bad Godesberg.
- Sozialdemokratische Partei Deutschlands (1969). *Regierungsprogramm der Sozialdemokratischen Partei Deutschlands 1969*. Bad Godesberg.
- Sozialdemokratische Partei Deutschlands (1972). *Wahlprogramm der SPD*. Bad Godesberg.
- Sozialdemokratische Partei Deutschlands (1976). SPD. *Weiter arbeiten am Modell Deutschland. Regierungsprogramm 1976-1980*. Bad Godesberg.
- Stachowiak, Herbert (1983). *Modelle – Konstruktionen der Wirklichkeit*. München.
- Statistisches Bundesamt. *Statistisches Jahrbuch* 1976, 1980, 1985, 1990, 1995, 1998. Wiesbaden.
- Steinmetz, A. (1983). The Discrepancy Evaluation Model. In G. F. Madaus (Hrsg.), *Evaluation Models. Viewpoints on Educational and Human Services Evaluation*, S. 79-99, Boston.

- Stufflebeam, D. L. (1972). Evaluation als Entscheidungshilfe. In C. Wulf (Hrsg.), *Evaluation. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen*, S. 113-145, München.
- Stufflebeam, D. L. & Webster, W. S. (1981). *An Analysis of Alternative Approaches to Evaluation*. Boston.
- Stufflebeam, D. L. & Shinkfield, A. J. (1985). *Systematic evaluation. A self-instructional Guide to Theory and Practice*. Boston.
- Stufflebeam, D. L. & Shinkfield, A. J. (1990). *Systematic evaluation*. Boston.
- Stufflebeam, D. L. (2001). *Evaluation Models*. San Francisco.
- Suchman, E. (1967). *Evaluation Research*. New York.
- Thorndyke, R. L. (1971). *Educational Measurement*. Washington D.C.
- Trochim, W. M. K. (1985). Pattern Matching, Construct Validity and Conceptualization in Program Evaluation. *Evaluation Review*. H. 5, S. 575-604.
- Trochim, W. M. K. (1986). *Advances in Quasi-Experimental Design and Analyses. New Directions for Program Evaluation*. San Francisco.
- Tyler, R. W. (1950). *Basic principles of curriculum and instruction*. Chicago.
- Twardy, M. (Hrsg.) (1995). *Modellversuchsforschung als Berufsbildungsforschung*. Köln.
- Vught, F. A. v. (1989). Creating Innovation in Higher Education. *European Journal of Education*. 24, S. 249-270, Ebington.
- Weber, B. (1992). *Innovation der Berufsausbildung durch Modellversuche? Kritische Anmerkungen zu veröffentlichten Modellversuchen*. Unveröffentlichte Hausarbeit im Rahmen der Ersten Staatsprüfung für das Lehramt an berufsbildenden Schulen im Lande Niedersachsen. Oldenburg.
- Weishaupt, H. (1980). Modellversuche im Bildungswesen und ihre wissenschaftliche Begleitung. In Max-Planck-Institut für Bildungsforschung (Hrsg.), *Bildung in der Bundesrepublik Deutschland*, S. 1287-1342, Hamburg.
- Weishaupt, H. (1983). Die finanziellen Ressourcen der Bildungsforschung. In: *Zeitschrift für erziehungs- und sozialwissenschaftliche Forschung*. 2, H. 1, S. 81-112.
- Weishaupt, H. (1991). *Begleitforschung zu Modellversuchen im Bildungswesen*. Frankfurt am Main.
- Weiss, C. H. (1972). *Evaluating Action Programs. Readings in Social Action and Education*. Boston.
- Weiss, C. H. (1995, 1997, 1998). *Evaluation*. New Jersey.
- Weiss, C. H. (1997). How can theory-based evaluation make greater headway? *Evaluation Review*, 21, S. 501-524.
- Weitz, Bernd. O. (1996). Qualitativ orientierte Modellversuchsforschung am Beispiel der Einzelfallstudie. Möglichkeiten und Grenzen qualitativ orientierter Bildungsforschung. In K.-H. Sommer (Hrsg.), *Bildungsforschung, Modellversuche und berufspädagogische Projekte*, S. 123-141. Suttgart.

- Westermann, R. (2002). Merkmale und Varianten von Evaluationen: Überblick und Klassifikation. *Zeitschrift für Psychologie*, H 210, S. 4-26.
- Westkamp, H. (2000). Innovationen optimieren – Inhaltliche Gestaltung und Außenwirkung von BLK-Programmen. In: D. de Haan (Hrsg.), *Bildung ohne Systemzwänge*, S. 291-298, Neuwied.
- Widmer, T. (1995, 1996). *Meta-Evaluation – Kriterien und Bewertung von Evaluation*. Bern.
- Widmer, T. (2000). Qualität von Evaluation – Wenn Wissenschaft zur praktischen Kunst wird. In R. Stockmann (Hrsg.), *Evaluationsforschung – Grundlagen und ausgewählte Forschungsfelder*, S. 77-103, Opladen.
- Widmer, T. & Binder, H.-M. (1997). Forschungsmethoden. In W. Bussmann (Hrsg.), *Einführung in die Politikevaluation*, S. 185-213, Basel.
- Winnefeld, F. (1972). Pädagogisches Feld als Faktorenkomplexion. In G. Dohmen (Hrsg.), *Unterrichtsforschung und didaktische Theorie*, S. 35-39, München.
- Wolf, R. L. (1973). The Use of Judicial Evaluation Methods in the Formulation of Educational Policy. In E. R. House (Hrsg.), *School evaluation – The politics and process*, S.198-203, Berkley.
- Wortmann, P. M. (1983). Evaluation Research: A Methodological Perspective. *Annual Review of Psychology*, 34, S. 223-260.
- Wottawa, H. & Thierau, H. (1998). *Lehrbauch Evaluation*. Bern.
- Wulf, C. (1972). *Evaluation. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen*. München.
- Wulf, C. (1977). *Theorie und Konzepte der Erziehungswissenschaft*. München.
- Zimmer, G. (1995). Schlussfolgerungen für die Modellversuchsreihe 'Multimediales Lernen in der Berufsbildung'. In H. Holz (Hrsg.), *Was leisten Modellversuche?* S. 277-294, Berlin.
- Zimmer, G. (1997). Wissenschaftliche Begleitung von Modellversuchen: Auf der Suche nach der Theorie innovativer Handlungen. *Berufsbildung in ,Wissenschaft und Praxis*, 26, H.1, S. 27-33.

**Zur Leistungsfähigkeit von Modellversuchen im kaufmännischen
Schulwesen der Bundesrepublik Deutschland –
eine metaevaluative Studie als Beitrag zur Modellversuchsarbeit**

Anhang zu Teil A der Studie:

- **Untersuchung der Berufsbildungspolitik nach Bundesministern**

A 1 Untersuchung der Berufsbildungspolitik nach Bundesministern

A 1.1 Untersuchung der Berufsbildungspolitik des Zeitraumes 1983 - 1987 (Willms)

Die Untersuchung der von Willms verfolgten berufsbildungspolitischen Intentionen ergibt eine Unterteilung in acht Zielbereiche. Die Zahlen, die in Klammern genannt sind, stellen die Anzahl der Nennungen innerhalb der jeweiligen Schwerpunktgruppe dar:

Stärkung des Ausbildungsplatzangebotes (3), Abbau/Veränderung von Rechtsvorschriften (5), Problemgruppen (3), Weiterbildung (5), Zusammenarbeit Bund und Länder (3), neue Technologien (3), Durchlässigkeit und Gleichwertigkeit (3) sowie Differenzierung/ Begabtenförderung (2).

Die Bundesregierung/Das BMBF [...]

1. will eine Stärkung des Ausbildungsplatzangebotes erreichen. Hierzu sollen a) das Ausbildungsplatzangebot des Staates erhöht, b) betriebliche Ausbildungskooperationen unterstützt, c) Sondermaßnahmen bedarfsgerecht gestaltet und d) hemmende Rechtsvorschriften abgebaut werden (Punkt 1 und 2, s. o., sind zusammengefasst) (vgl. BMBF 11/1982: 197ff; 6/1983: 101f; 2/1985: 45 sowie Bundesregierung 121/1983: 1103ff; 85/1984: 753ff; 88/1985: 770; 84/1985: 732ff; 135/1986: 1135f).
2. will die "Problemgruppen" (Jugendliche ohne Schulabschluss, Behinderte, ausländische Jugendliche und Frauen) bei ihrem Berufseinstieg unterstützen und Berufsperspektiven ausbauen (vgl. BMBF 11/1982: 197f; 6/1983: 101f; 2/1985: 45 sowie Bundesregierung 121/1983: 1103ff).
3. will die Weiterbildungsbereitschaft der Arbeitnehmer fördern. Dazu sollen sich Inhalte stärker am Bedarf der Wirtschaft orientieren, Weiterbildungseinrichtungen modernisiert, die Zusammenarbeit von Betrieben und beruflichen Schulen, Fachhochschulen und Universitäten gefördert und die fachlichen und pädagogischen Qualifikationen der Ausbilder ausgebaut werden (vgl. BMBF 11/1982: 197f; 7-8/1986: 100 sowie Bundesregierung 17/1983: 160).
4. will durch die stärkere Zusammenarbeit von Bund und Ländern die Situation der Berufsschule verbessern, die Bekämpfung der Schwierigkeiten des Bildungs- und Beschäftigungssystems intensivieren und das quantitative Angebot beruflicher Vollzeitschulen erhalten. (vgl. BMBF 11/1982: 197f; 9/1984: 152; 2/1985: 27 sowie Bundesregierung 85/1984: 753ff; 88/1985: 770).
5. verfolgt das Ziel die ‚neuen Technologien‘ verstärkt zum Einsatz zu bringen. Hierzu wird das Aktionsprogramm neue Technologien initiiert und eine entsprechende Modellversuchsreihe ins Leben gerufen (vgl. BMBF 11/1982: 197f; 2/1985: 27 sowie Bundesregierung 88/1985: 770).
6. will die Anerkennung beruflicher und allgemeiner Bildung als gleichwertig durch neue Regelungen beim Hochschulzugang unterstützen (vgl. BMBF 11/1982: 197f; 9/1984: 152 sowie Bundesregierung 15/1983: 141ff).

7. protegiert die Differenzierung des Ausbildungsplatzangebotes und der Unterrichtsgestaltung um lernschwächere und lernstärkere Jugendliche zu motivieren (vgl. BMBF 11/1982: 197f; 2/1985: 124 sowie Bundesregierung 85/1984: 753ff).

Im Hinblick auf die Stärkung des Ausbildungsplatzangebotes verzeichnet das BMBF jährliche Zuwachsraten (vgl. BMBF 1983-1987), die auf verschiedene Aktivitäten zurückzuführen sind. So sind die Ausbildungskapazitäten des öffentlichen Dienstes kontinuierlich gesteigert worden (vgl. BMBF 1984: 1ff); Gespräche mit Betrieben, Kammern, Arbeitnehmer- und Arbeitgebervertretern führten zu einer Sensibilisierung für diesen Problembereich; hier konnten dann u. a. Möglichkeiten wie Ausbildungsverbundmodelle zur Kooperation im Bereich der Zusammenlegung von Teilausbildungskapazitäten geschaffen werden (vgl. BMBF 1985: 74). Auch die Modernisierung verschiedener Ausbildungsordnungen führten zu einem weiteren Annähern an die aktuelle Berufsbildungspraxis und soll zu einer Stärkung der Ausbildungsbereitschaft beigetragen haben (vgl. BMBF 1983-1987).

Bei der Förderung der sogenannten Problemgruppen sind nach Angaben des BMBF, bezogen auf dessen Dokumentation 'Berufsbildungsbericht' Erfolge in verschiedenen Schwerpunktbereichen zu verzeichnen. Neben der Entwicklung gesondert geregelter Ausbildungsgänge für Behinderte (vgl. BMBF 1983: 59), den ausbildungsbegleitenden Hilfen für Jugendliche mit schlechten schulischen Voraussetzungen oder sozialen Benachteiligungen (vgl. BMBF 1983: 63; 1984: 77; 1985: 80 u. v. a.), den Maßnahmen zur Berufsvorbereitung und sozialer Eingliederung junger Ausländer (MBSE) (BMBF 1983: 67; 1985: 86) ist vor allem das Programm für die Förderung der Berufsausbildung von benachteiligten Jugendlichen zu erwähnen, das jährlich mit Beträgen in Millionenhöhe durch den Bund gefördert wurde (vgl. BMBF 1983-1987). Daneben gab es außerdem Modellprogramme für junge Frauen, die u. a. die Eingliederung in die gewerblich-technische Ausbildung, in sog. 'Männerberufe', zum Ziel hatten oder das Ausländer Modellprogramm zur Integration und sprachlichen Förderung junger Ausländer (vgl. BMBF 1983:56, 1984: 85).

In Bezug auf den Bereich Weiterbildung sind vor allem die verschiedenen Maßnahmen hervorzuheben, die durch das Arbeitsförderungsgesetz (AFG) auf eine finanzielle Grundlage gestellt wurden und deren Leistungen 1987 durch das Institut für Arbeitsmarkt- und Berufsforschung evaluiert wurden. Ein Schwerpunkt hier waren Weiterbildungsmaßnahmen, die als Eingliederungsmaßnahmen initiiert wurden (vgl. BMBF 1985: 91). In diesem Zusammenhang sei auch die Modellreihe für Erwachsene erwähnt, die keine abgeschlossene Berufsausbildung vorweisen können (BMBF 1985: 95). Außerdem dokumentiert das Ministerium eine steigende Anzahl von Fortbildungsprüfungen (BMBF 1986: 97) sowie die Modellreihe 'Erschließung gewerblich-technischer Ausbildungsberufe für Mädchen', deren hohe Erfolgsquote bei den Abschlussprüfungen und die gute Eingliederung der Teilnehmerinnen nach dem Modellvorhaben als Erfolg gewertet wurde (vgl. BMBF 1986: 79). 1987 wurde dann die Qualifizierungsoffensive des Bundes in Zusammenarbeit mit den Sozial-

partnern gestartet und neue Fortbildungsverordnungen erlassen (vgl. BMBF 1987: 126ff).

Die Förderung der 'neuen Technologien' wurde u. a. durch die Anpassung der Ausbildungsordnungen an die technischen Entwicklungen, insbesondere der Mikroelektronik zu Beginn der 80-er Jahre (BMBF 1983: 54) sowie durch die Modellversuchsreihe 'neue Technologien' 1984 (BMBF 1984: 77) unterstützt. Hier stand vor allem auch die Aufarbeitung von fachlichen Defiziten und didaktisch-methodischen Mängeln im Vordergrund (BMBF 1985: 71). Für 1987 hat man sich erste Ergebnisse versprochen (BMBF 1987: 102).

Der Aspekt ‚Zusammenarbeit zwischen Bund und Ländern‘ wurde in den Berufsbildungsberichten dieses Betrachtungszeitraumes nicht explizit hervorgehoben. Dennoch ist offensichtlich, dass viele bildungspolitische Ziele nur in Zusammenarbeit mit den Ländern verfolgt werden konnten, da der Bereich Bildung und Schule unter den Grundsatz der Länderhoheit fällt.

Die Zielbereiche 'Durchlässigkeit und Gleichwertigkeit' sowie 'Differenzierung und Eliteförderung' können zusammenfassend wie folgt dargestellt werden: Gleichwertigkeit beruflicher und allgemeiner Bildung wird vordringlich unter dem Aspekt des Berechtigungswesens betrachtet und durch die Gleichstellung von Berufsabschlüssen mit mittleren allgemeinbildenden Abschlüssen unterstützt (BMBF 1984: 7). Des Weiteren soll die Möglichkeit des Fachhochschulzugangs für Meister erreicht werden (BMBF 1985: 2). Differenzierung und Begabtenförderung werden in dieser Amtszeit überwiegend unter der Maßgabe von Fortbildungs- und Fernlehrangeboten betrachtet (BMBF 1983: 76; 1984: 7; 1986: 81).

A 1.2 Untersuchung der Berufsbildungspolitik des Zeitraumes 1987 – 1991 (Möllemann)

Die vom Bundesbildungsminister Möllemann verfolgten berufsbildungspolitischen Intentionen ergeben eine Unterteilung in sechs Zielbereiche. Die Zahlen, die in Klammern genannt sind, stellen die Anzahl der Nennungen innerhalb der jeweiligen Schwerpunktgruppe dar: Stärkung des Ausbildungsplatzangebotes (3), Abbau/Veränderungen von Rechtsvorschriften, (1), Problemgruppen (7), Aus- und Fortbildung (8), Neue Technologien (1), Durchlässigkeit und Gleichwertigkeit (1).

Die Bundesregierung/Das BMBF [...]

1. will eine Stärkung des Ausbildungsplatzangebotes durch eine Intensivierung der Kooperation der Lernorte Betrieb und Schule, durch eine Lehrstellenoffensive und durch eine Verbesserung der Ausbildungsplatzförderung erreichen (vgl. BMBF 3/1987: 92f; 2/1990: 10f sowie Bundesregierung 33/1987: 269ff).
2. will zu einer Modernisierung von Ausbildungs- und Fortbildungsordnungen beitragen (vgl. BMBF 5/1987: 148f).
3. will zur Eingliederung von Benachteiligten bzw. Mitgliedern sog. 'Problemgruppen' beitragen, indem a) die Möglichkeit von Teilqualifikationen im Sin-

- ne einer modularen Ausbildung angeboten wird, b) Modellversuche für junge Frauen und Jugendliche initiiert werden, c) das bildungspolitische Gesamtkonzept durch entsprechende Angebote erweitert wird (vgl. BMBF 3/1987: 92; 5/1987: 148ff sowie Bundesregierung 33/1987: 269ff; 141/1987: 1198f; 29/1989: 261f; 44/1989: 393f).
4. will a) die Qualifizierung von Ausbildern bzw. Mitgliedern von Berufsbildungs- und Prüfungsausschüssen fördern, b) die Teilnahme an Weiterbildungsmöglichkeiten stärken, c) Weiterbildung als vierte Säule (quartärer Sektor) etablieren und somit auch die Attraktivität des dualen System steigern und d) die Weiterbildungsangebote für Ungelernte ausbauen (vgl. BMBF 6/1987: 181f; 3/1988: 39; 2/1990: 10 sowie Bundesregierung 141/1987: 1198; 39/1988: 328ff; 122/1988: 1113ff; 26/1989: 261f; 4/1989: 393f).
 5. will die Akzeptanz und die Verbreitung neuer Technologien durch Modellversuche und durch Erziehung im Umgang mit den neuen Technologien fördern (vgl. BMBF 6/1987: 182).
 6. will die Gleichwertigkeit beruflicher und allgemeiner Bildung durch die Gleichstellung beruflicher und schulischer Abschlüsse unterstützen (vgl. BMBF 3/1988: 36f)

Das Bundesbildungsministerium stellt in der gesamten Amtsperiode fest, dass das Ausbildungsplatzangebot die Ausbildungsplatznachfrage übersteigt. Dabei wurden strukturelle und regionale Ungleichheiten nicht berücksichtigt, d. h. der bestehende Nachfrageüberhang für Ausbildungsberufe im Handel und im kaufmännischen Bereich bzw. Unterschiede in Nord und Süd sowie Ost und West wurden bei der Beschreibung des Ausbildungsplatzangebotes nicht mit einbezogen (BMBF 1987 - 1991). Um dem Problem des regionalen und strukturellen Missverhältnisses entgegenzuwirken hat der Bund 1990 ein Sonderprogramm initiiert, das der Verbesserung der regionalen Wirtschaftsstruktur dienen sollte. Darüber hinaus wurden überbetriebliche berufliche Ausbildungsstätten gefördert (BMBF 1990: 1f).

Auch im Zeitraum 1988 bis 1991 wurden die bestehenden Ausbildungsordnungen kontinuierlich erneuert und für gültig erklärt. Des Weiteren konnte jetzt damit begonnen werden, diejenigen Ausbildungsordnungen zu überarbeiten, die bereits die nach dem BBiG geordneten Berufe betrafen.

Ein Schwerpunkt im Themenbereich 'Problemgruppen' war die Arbeit zur Erstellung von Programmen für die Förderung der Berufsausbildung von benachteiligten Jugendlichen. Das Spektrum reichte hier von der Möglichkeit ausbildungsbegleitender Hilfen (ABH), über regionale Projekte, die eine intensive Informationsarbeit zum Berufswahlspektrum zum Gegenstand hatten (BMBF 1988: 90) bis hin zu spezifischen Modellversuchen (BMBF 1990: 126). Auch die Initiative 'BüE', Berufsausbildung in einer überbetrieblichen Einrichtung, ist zu diesem Bereich hinzuzuzählen (BMBF 1991: 97).

Die 1988 vom Bund gestartete Qualifizierungsinitiative diente u. a. der Wiedereingliederung von Un- und Angelernten sowie Langzeitarbeitslosen (BMBF 1988: 111f) und ist der berufsbildungspolitischen Zielsetzung 'Weiterbildung' zuzuordnen. Darüber hinaus gab es verschiedene Modellvorhaben, die das Bildungsverhalten unterstützen sollten bzw. der Wiedereingliederung von Frauen nach der Kinderpause dienten. Auch Modellversuche zur pädagogischen Beratung im Handwerk wurden ins Leben gerufen (BMBF 1988: 130; 1990: 162). Am Ende der Amtsperiode wurden dann auch schon erste Ergebnisse aus der konzertierten Aktion Weiterbildung erwartet, die noch von der Amtsvorgängerin, Frau Willms, gestartet wurde.

Die Auseinandersetzung mit den 'Neuen Technologien' erfolgte vor allem durch die Zusammenfassung von Einzelvorhaben zu Modellversuchsreihen um hier übergreifende Aussagen treffen zu können (BMBF 1988: 2). Häufig erfolgte keine isolierte Darstellung dieses Themenbereiches, sondern es wurden Ergebnisse in den Bereichen Weiterbildung oder Ausbildung kurz dokumentiert.

Zum Themenaspekt 'Gleichwertigkeit beruflicher Bildung', der als berufsbildungspolitische Zielintention definiert wurde, erfolgte keine Dokumentation von Resultaten oder Teilerfolgen. Es wurde außerdem nicht deutlich, ob dieser Zielbereich evtl. durch die berufliche Weiterbildung unterstützt werden sollte oder ob Aufstiegsfortbildung für begabte Jugendliche geplant war.

A 1.3 Untersuchung der Berufsbildungspolitik des Zeitraumes 1991 - 1994 (Ortleb)

Die Untersuchung der von Herrn Ortleb verfolgten berufsbildungspolitischen Intentionen ergibt eine Unterteilung in sechs Zielbereiche. Die Zahlen, die in Klammern genannt sind, stellen die Anzahl der Nennungen innerhalb der jeweiligen Schwerpunktgruppe dar:

Stärkung des Ausbildungsplatzangebotes (2), Differenzierung/Elite (1), Durchlässigkeit und Gleichwertigkeit (2), Europa (2), Aufbau Ost (5), Attraktivitätssteigerung des dualen Systems (2).

Die Bundesregierung/Das BMBF [...]

1. will eine Stärkung des Ausbildungsplatzangebotes durch eine Ausbildungs-offensive Ost sowie durch die gegenseitige Anerkennung von Schulabschlüssen fördern (vgl. BMBF 3/1991: 27; 2/1991).
2. will durch Differenzierungen im Schulformenangebot eine Förderung von lernschwachen und leistungsfähigen Schülern vornehmen (vgl. BMBF 3/1991: 27).
3. will die Durchlässigkeit innerhalb des Bildungssystems bzw. Gleichwertigkeit beruflicher Bildung durch die Förderung der Aufstiegsfortbildung zum Meister und durch den Hochschulzugang nach der Berufsausbildung unterstützen (vgl. BMBF 3/1991: 28; 4/1991).
4. will den europäischen Gemeinschaftsgedanken durch eine gemeinsame Bildungspolitik und durch Eingliederungsmöglichkeiten in den jeweiligen nationalen Arbeitsmarkt in das Bildungssystem tragen (vgl. BMBF 2/1992: 9;

- 12/1992: 161).
5. verfolgt das Ziel, den Aufbau des Bildungssystems der neuen Bundesländer durch a) den Ausbau überbetrieblicher Bildungsanstalten, b) den Ausbau der Berufsschule, c) die Qualifizierung der Ausbilder, d) die Förderung der Weiterbildung und durch e) den Ausbau des Fremdsprachenunterrichts voran zu treiben (vgl. BMBF 3/1991: 28; 2/1992: 9f sowie Bundesregierung 42/1991: 312; 118/1992: 1085).
 6. will eine Attraktivitätssteigerung des dualen Systems durch a) Aufstiegsfortbildungen für Abiturienten, b) strukturelle Veränderungen vornehmen (vgl. BMBF 2/1992: 58f sowie Bundesregierung 42/1991: 312).

Die Ausbildungsinitiative Ost, die sich die Lösung des Problems des Nachfrageüberhangs in Bezug auf Ausbildungsplätze in den neuen Bundesländern zur Aufgabe gemacht hatte, strebte dieses Ziel vor allem durch die Schaffung außerbetrieblicher Ausbildungsplätze an (BMBF 1992: 82; 1993: 9). Ab 1994 wurde dieses Vorhaben durch eine Gemeinschaftsinitiative von Bund, Ländern, der Stadt Berlin sowie dem Europäischen Sozialfonds mit insgesamt 500 Millionen DM unterstützt. Hier konnten dann bis zu 10.000 Ausbildungsplätze zusätzlich geschaffen werden, wovon wiederum 70 % im begehrten Bereich der kaufmännischen Berufe und der Dienstleistungsberufe lagen (BMBF 1994: 89f).

Der Aspekt der Differenzierung wurde im Rahmen dieser Amtszeit unter dem Blickwinkel der formalen und der pädagogischen Differenzierung betrachtet. So wurde mit dem § 40 c AFG die Möglichkeit geschaffen, lernschwache Jugendliche bei der Schließung von Ausbildungslücken zu unterstützen (BMBF 1992: 94). Darüber hinaus ist ein BLK Programm aufgenommen worden, das sich der Differenzierung in der Berufsbildung widmete (BMBF 1993: 4). 1994 dann wurde ein Handlungskonzept initiiert, das ‚differenzierte Wege zum anerkannten Berufsabschluss‘ hervorbringen sollte und somit als Prävention von Arbeitslosigkeit angesehen wurde (BMBF 1994: 91).

Die Förderung leistungsstarker Jugendlicher wurde u. a. durch ein Programm zur Begabtenförderung für etwa 3000 Personen (BMBF 1992: 8) vorgenommen. Weiterhin konnten neue Ausbildungsformen wie die ‚Betriebsassistenten im Handel‘ geschaffen werden (BMBF 1993: 5). Der Aspekt der Weiterbildung ist außerdem unter diesen Gesichtspunkt zu subsumieren: Hier sind es verschiedenartige Stipendien, die in einer Gesamtzahl von 8.100 Plätzen in einer Laufzeit von drei Jahren monatlich 3.000,00 DM zur Verfügung stellten. Hierdurch sollte die Attraktivität des dualen Systems erheblich gesteigert werden (BMBF 1994: 94).

Die Punkte Durchlässigkeit und Gleichwertigkeit beruflicher Bildung wurden durch die Gleichsetzung des Berufsabschlusses mit dem mittleren allgemeinbildenden Abschluss und eines guten Meisterzeugnisses mit dem Zugang zu einem fachgebundenen Hochschulstudium diskutiert (BMBF 1993: 6), jedoch hier noch nicht umgesetzt.

Die Förderung des europäischen Gemeinschaftsgedanken hat der Minister durch die Anerkennung beruflicher Befähigungsnachweise sowie durch ein Sonderprogramm zur Ausbildung von Sprachkenntnissen (BMBF 1992: 169) zu unterstützen versucht. Darüber hinaus hat die Bundesregierung auf das Memorandum der EG-Kommission über die Berufsbildungspolitik der 90-er Jahre erwidert, dass sie Transparenz bei der Form der Ausbildungsgestaltung und der Befähigungsnachweise fordere (BMBF 1994: 150).

Der Aufbau des Berufsbildungssystems Ost wurde vor allem durch die AFG-Förderungen in Form von Zuschüssen für Bau, Erweiterung und Ausstattung von außer- und überbetrieblichen Lehrwerkstätten unterstützt. Außerdem hat die Bundesanstalt für Arbeit Darlehen für Fortbildungen, Umschulungen und für die Einarbeitung in neue Arbeitstätigkeiten gewährt (BMBF 1994: 107.) 1994 wurde dann auch das 'Programm zur Qualifizierung von Personal der beruflichen Bildung in den neuen Ländern' (PQO) ins Leben berufen. Hier geht es vor allem um den Transfer von Erfahrungen und Ergebnissen aus den alten in die neuen Bundesländer (BMBF 1994: 98).

Die Attraktivität des dualen Systems sollte u. a. durch die Initiierung von Sonderausbildungslehrgängen für Abiturienten sowie durch Maßnahmen zur Differenzierung gesteigert werden (BMBF 1992: 103; 1994: 94). Darüber hinaus sollten Berufsakademien gegründet werden um die Möglichkeit des Einstiegs in mittlere Führungsebenen auch durch eine Berufsausbildung zu ermöglichen (BMBF 1992: 103).

A 1.4 Untersuchung der Berufsbildungspolitik des Zeitraumes 02/1994 - 11/1994 (Laermann)

Die Untersuchung der von Herrn Laermann verfolgten berufsbildungspolitischen Intentionen ergibt eine Unterteilung in fünf Zielbereiche. Die Zahlen, die in Klammern genannt sind, stellen die Anzahl der Nennungen innerhalb der jeweiligen Schwerpunktgruppe dar:

Abbau/Veränderung von Rechtsvorschriften (1), Differenzierung/Elite (1), Durchlässigkeit und Gleichwertigkeit (1), Europa (3), Attraktivitätssteigerung des dualen Systems (4).

Die Bundesregierung/Das BMBF [...]

1. will Änderungen in den Ausbildungsordnungen vorantreiben (vgl. BMBF 1-2/1994).
2. will erreichen, dass das Angebot für ausbildungsbegleitende und zertifizierbare Maßnahmen ausgebaut wird (vgl. BMBF 3/1994: 16f).
3. will das Angebot verschiedener Möglichkeiten, die zum Hochschulzugang führen, ausbauen (vgl. BMBF 3/1994: 18f).
4. will internationale Austauschmaßnahmen und die Ausweitung der Fremdsprachenkenntnisse in der beruflichen Bildung sowie die Vermittlung von Europakenntnissen vorantreiben (vgl. BMBF 3/1994: 18; 9/1994: 99 sowie Bundesregierung 35/1994: 315f).
5. will die Attraktivität des dualen Systems durch a) die Entwicklung neuer in-

haltlich und zeitlich zusammenhängender Ausbildungswege für mittlere Führungspositionen, b) den Ausbau von Berufsakademien und die Kombination von Fachhochschulstudium und betrieblicher Ausbildung und c) den Ausbau von Informations- und Beratungsdiensten im Hinblick auf Aufstiegs- und Weiterbildungsmöglichkeiten steigern (vgl. BMBF 3/1994: 18, 20 sowie Bundesregierung 35/1994: 315f; 47/1994: 119f).

Auf Grund der kurzen Amtszeit von Bundesminister Laermann sind Aussagen zum Zielerreichungsgrad nicht möglich, da die Aussagen des Berufsbildungsberichtes 1995 nicht eindeutig den Bemühungen Ortlebs oder Laermanns zugeordnet werden können. Die Explikation der Zielbereiche kann aber dazu dienen, Entwicklungslinien, wenn in diesem Bereich auch nur ansatzweise, nachzuvollziehen.

A 1.5 Untersuchung der Berufsbildungspolitik des Zeitraumes 1994 - 1998 (Rüttgers)

Die Untersuchung der von Herrn Rüttgers verfolgten berufsbildungspolitischen Intentionen stellt sich als schwierig dar, da der Bundesbildungsminister auf Zielformulierungen bzw. auf die Veröffentlichung eines Strukturprogramms verzichtet hat. Hier sind es lediglich Schlagworte, die aus wenigen öffentlichen Stellungnahmen entnommen werden konnten. Diese beziehen sich auf die Aspekte Differenzierung von Bildungsangeboten, Vermittlung von Lerntechniken und Begabtenförderung. Der jeweilige Teil I der Berufsbildungsberichte kann jedoch in Ermangelung anderer Äußerungen als Stellungnahme des Bundesministeriums zu den Schwerpunkten der Berufsbildungspolitik angesehen werden. Die Aussagen, die hieraus gewonnen werden können, beziehen sich dann nur auf tatsächliche Leistungen und können herangezogen werden um Entwicklungslinien der Berufsbildungspolitik nachzuzeichnen.

Wie in den vorhergehenden Amtsperioden ist auch hier die Ausbildungsplatzsituation ein Schwerpunkt der laufenden Berufsbildungspolitik. Während in den alten Bundesländern ein Angebotsüberhang dokumentiert wird, musste die Bundesregierung in den neuen Bundesländern das eigene Angebot an Ausbildungsplätzen wieder ausweiten. Darüber hinaus sind außerbetriebliche Ausbildungsplätze weiter gefördert worden (BMBF 1995: 2; 1996: 2). In diesen Zusammenhang ist auch das Aktionsprogramm ‚Lehrstellen Ost‘ einzuordnen, das mit 190 Millionen DM 1996 initiiert wurde (BMBF 1997: 1). Erst 1997 stellt das Ministerium eine Zunahme um 2 % der betrieblichen Ausbildungsstellen in den neuen Bundesländern fest.

Die Attraktivität des dualen Systems will das Bundesministerium durch die permanente Aktualisierung der Ausbildungsordnungen gesteigert gesehen haben (BMBF 1995: 7). In diesem Zusammenhang sind auch die Bemühungen um die Gleichwertigkeit allgemeiner und beruflicher Bildung einzuordnen. Die öffentliche Verwaltung bemühte sich neue berufliche Entwicklungsmöglichkeiten aufzuzeigen und flexible und leistungsgerechte Entgeltstrukturen zu entwickeln (BMBF 1997: 8).

Ein weiterer Schwerpunkt der politischen Tätigkeit in dieser Amtsperiode scheint die Orientierung an der internationalen, vor allem der europäischen Bildungspolitik zu sein. Hier sind vor allem die EU Bildungsprogramme Sokrates und Leonardo zu nennen, aber auch die Förderung von Austauschmaßnahmen junger Auszubildender (BMBF 1996: 19, 139). Außerdem ist eine aktive berufsbildungspolitische Zusammenarbeit, die eine Ausrichtung auf die betriebliche Praxis erfahren soll, wie z. B. die Entwicklung von mehrsprachigen Befähigungsnachweisen (BMBF 1997: 9), vorgesehen.

**Zur Leistungsfähigkeit von Modellversuchen im kaufmännischen
Schulwesen der Bundesrepublik Deutschland –
eine metaevaluative Studie als Beitrag zur Modellversuchsarbeit**

Anhang zu Teil B der Studie:

- **Evaluationsraster auf Basis elaborierter Evaluationsstandards**
- **Ausgewählte Modellversuche der Gruppe 4: 'Innovationen in bestehenden Fächern oder Schulformen'**

N1 - Ermittlung der Beteiligten und Betroffenen

Definition des Standards:

Die an einer Evaluation beteiligten oder von ihr betroffenen Personen sollten identifiziert werden, damit deren Interessen und Bedürfnisse berücksichtigt werden.

Operationalisierung:

Die B&B¹ werden, sofern es sich um potenziell wichtige Personen handelt und die zeitlichen und finanziellen Beschränkungen dies zulassen, identifiziert, ihre Informationsbedürfnisse ermittelt und in die Planung und Durchführung mit einbezogen. Weniger einflussreichen Gruppen wird eine ausreichende Beteiligung ermöglicht.

Ausprägungsgrad 0:

Eine Identifikation der B&B findet nicht statt.

Ausprägungsgrad 1:

Die Identifikation und Gruppierung der B&B findet nicht nach dem Kriterium der Einflussmöglichkeit statt. Eine Einbeziehung der Personengruppen unterbleibt.

Ausprägungsgrad 2

B&B werden identifiziert und nach ihrer Einflussmöglichkeit gruppiert. Lediglich die Informationsbedürfnisse der potenziell wichtigen B&B werden berücksichtigt bzw. diese werden hinsichtlich Planung und Durchführung mit einbezogen.

Ausprägungsgrad 3

Die B&B werden hinsichtlich des Ausmaßes ihrer Einflussmöglichkeiten auf den Modellversuch gruppiert und entsprechend (stärker oder weniger stark) in die Planung und Durchführung mit einbezogen.

N3 – Umfang und Auswahl der Informationen

Definition des Standards:

Die gewonnenen Informationen sollten von einem Umfang und einer Auswahl sein, welche die Behandlung sachdienlicher Fragen zum Modellversuch ermöglichen und gleichzeitig auf die Interessen und Bedürfnisse des Auftraggebers und anderer B&B eingehen.

Operationalisierung:

Die Evaluatoren bestimmen das Informationspotenzial und selektieren diejenigen Informationen, die im Sinne des Modellversuchs wichtig sind und sondern sie somit aus der Informationsflut aus. Darüber hinaus werden auch Informationen ausgewählt, deren Bedeutung von den Auftraggebern bislang übersehen wurde.

Ausprägungsgrad 0:

Das Ermitteln der Info-Bedürfnisse aus dem MV heraus unterbleibt. Ein Herausheben bedeutender Infos im Sinne des MV oder weitergehender Interessen aus der Informationsflut kann somit nicht erfolgen.

Ausprägungsgrad 1:

Die Evaluatoren ermitteln die für den Auftraggeber relevanten Info-Bedürfnisse. Allerdings werden die bedeutenden Infos nicht aus der Informationsmenge selektiert oder weitergehende Informationsbedürfnisse bleiben unberücksichtigt.

1) B&B=Beteiligte und Betroffene

Ausprägungsgrad 2:

Die Evaluatoren ermitteln die für den Auftraggeber relevanten Info-Bedürfnisse und selektieren die bedeutenden Informationen. Weitergehende Details werden nicht hervorgehoben.

Ausprägungsgrad 3:

Die Evaluatoren ermitteln den Informationsbedarf des MV, diese Infos werden selektiert. Darüber hinaus werden vom Auftraggeber und/oder MV bislang nicht als bedeutend erachtete Infos hervorgehoben.

N4 – Feststellung von Werten**Definition des Standards:**

Die Perspektiven, Verfahren und Gedankengänge, auf denen die Interpretationen der (Evaluations-; Anm. Verf.) Ergebnisse beruhen, sollten sorgfältig beschrieben werden, damit die Grundlage der Werturteile klar ersichtlich ist.

Operationalisierung:

Die der Interpretation der Evaluationsergebnisse zugrunde gelegten Wertmaßstäbe werden offen gelegt und die Perspektiven, Verfahren, Gedankengänge werden beschreibend dokumentiert, bevor auf deren Grundlage Deutungen vorgenommen und Urteile ausgesprochen werden.

Ausprägungsgrad 0:

Perspektiven, Verfahren und Gedankengänge sowie die zu Grunde gelegten Wertmaßstäbe werden nicht dokumentiert; selbst Schlussfolgerungen und Urteile/Empfehlungen unterbleiben. Es erfolgt lediglich eine Verlaufsdokumentation.

Ausprägungsgrad 1:

Perspektiven, Verfahren und Gedankengänge werden dokumentiert; die zu Grunde gelegten Wertmaßstäbe bleiben jedoch verborgen.

Ausprägungsgrad 2:

Perspektiven, Verfahren und Gedankengänge werden dokumentiert; die zu Grunde gelegten Wertmaßstäbe auf die sich die Urteile und Empfehlungen gründen, bleiben jedoch unklar.

Ausprägungsgrad 3:

Perspektiven, Verfahren und Gedankengänge sowie die zu Grunde gelegten Wertmaßstäbe werden dokumentiert und darauf aufbauend Urteile und Empfehlungen ausgesprochen.

N5 – Klarheit des Berichts**Definition des Standards:**

Evaluationsberichte sollten den evaluierten Modellversuch einschließlich seines Kontextes ebenso beschreiben wie die Ziele, die Verfahren und Befunde der Evaluation, damit die wesentlichen Informationen zur Verfügung stehen und leicht verstanden werden können.

Operationalisierung:

Die B&B können leicht erkennen, welche Ziele die Evaluation verfolgte, was evaluiert wurde, wie die Evaluation durchgeführt wurde. Darüber hinaus ist ersichtlich, welche Informationen gewonnen wurden und zu welchen Schlussfolgerungen man kam und ggf. welche Empfehlungen gemacht wurden.

Ausprägungsgrad 0:

Die Ziele der Evaluation sowie Gegenstände, Informationen und Schlussfolgerungen werden nicht offen gelegt.

Ausprägungsgrad 1:

Die Ziele der Evaluation sowie die Evaluationsgegenstände werden bekannt gegeben.

Ausprägungsgrad 2:

Neben den Zielen werden auch die Evaluationsgegenstände und die gewonnenen Informationen genannt.

Ausprägungsgrad 3:

Die den Evaluatoren bekannten Daten (Ziele, Gegenstände, Verfahren, Informationen und Schlussfolgerungen) werden offen gelegt.

N – Nützlichkeitsstandards**Definition des Standards:**

Die Nützlichkeitsstandards sollen sicherstellen, dass sich eine Evaluation an den Informationsbedürfnissen der vorgesehenen Evaluationsnutzer ausrichtet.

Operationalisierung:

Die Evaluation ist derart aufgebaut, dass sie die Informationsbedürfnisse der B&B ermittelt und diese im Aufbau, der Durchführung und Auswertung auch berücksichtigt.

Ausprägungsgrad 0:

Die Modellversuchsevaluation richtet sich nicht an den Informationsbedürfnissen der B&B aus.

Ausprägungsgrad 1:

Die Modellversuchsevaluation berücksichtigt die Informationsbedürfnisse der B&B zu einem geringen Teil.

Ausprägungsgrad 2:

Die Modellversuchsevaluation berücksichtigt die Informationsbedürfnisse der B&B in befriedigendem Maße.

Ausprägungsgrad 3:

Die MV-Evaluation berücksichtigt die Informationsbedürfnisse der B&B.

D1 – Praktische Verfahren**Definition des Standards:**

Die Evaluationsverfahren sollten praktisch sein, so dass Störungen minimiert und die benötigten Informationen beschafft werden können.

Operationalisierung:

Die ausgewählten Evaluationsverfahren (Art der Evaluationsinstrumente, Einsatz und Informationssammlung) sind geeignet, die benötigten Informationen zu beschaffen; hinsichtlich ihres Aufwandes sind diese im Hinblick auf den zeitlichen Rahmen vertretbar.

Ausprägungsgrad 0:

Die ausgewählten Verfahren sind nicht vertretbar, d. h. diese greifen zu kurz bzw. überschreiten den Rahmen und/oder führen zu erheblichen Störungen.

Ausprägungsgrad 1:

Auswahl (Art), Einsatz und Auswertung der Evaluationsverfahren sind in Ansätzen geeignet, die benötigten Informationen zu erheben und entsprechen in etwa dem zeitlichen Rahmen und/oder führen zu noch vertretbaren Störungen.

Ausprägungsgrad 2:

Auswahl (Art), Einsatz und Auswertung der Evaluationsverfahren sind geeignet, einen Großteil der benötigten Informationen zu erheben und entsprechen in großem Maße dem zeitlichen Rahmen und/oder führen zu geringen Störungen.

Ausprägungsgrad 3:

Auswahl (Art), Einsatz und Auswertung sind angemessen; die Verfahren sind geeignet, die benötigten Informationen zu erheben und entsprechen dem zeitlichen Rahmen und führen nicht/kaum zu Störungen.

K1 – Unterstützung der Dienstleistungsorientierung**Definition des Standards:**

Die Evaluation sollte so geplant werden, dass Organisationen dabei unterstützt werden, die Interessen und Bedürfnisse des ganzen Zielgruppenspektrums zu berücksichtigen und ihre Tätigkeiten danach auszurichten.

Operationalisierung:

Die Auswirkungen eines Modellversuchs (Zielsetzungen) werden im Hinblick auf die Bedürfnisse (Mangelempfinden) der B&B untersucht. Die Evaluationsplaner blicken über die Interessen von Pädagogen und Organisationen hinaus und berücksichtigen auch die Entwicklung der Lernenden und der Gesellschaft. Die Ergebnisse der Evaluation sind geeignet, als Entscheidungshilfe für einen Planungsmodellversuch oder als Grundlage für bildungspolitische Entscheidungen herangezogen zu werden.

Ausprägungsgrad 0:

Die Auswirkungen des MV im Hinblick auf die Bedürfnisse der B&B werden nicht untersucht. Außerdem wird nicht über die Interessen der Pädagogen und Organisationen hinausgeblickt und so die Entwicklung der Lernenden nicht ausreichend berücksichtigt. Die Evaluation bringt demnach keine Ergebnisse hervor, die als Entscheidungshilfe für Planungsmodellversuche oder bildungspolitische Entscheidungen herangezogen werden können.

Ausprägungsgrad 1:

Den Auswirkungen des MV wird nachgegangen, allerdings werden keine Konsequenzen aus den Erkenntnissen gezogen. Die Evaluation bringt demnach nur bedingt Ergebnisse hervor, die als Entscheidungshilfe für Planungsmodellversuche oder bildungspolitische Entscheidungen herangezogen werden können.

Ausprägungsgrad 2:

Die Evaluationsplaner berücksichtigen die Entwicklung der Lernenden indem sie die Auswirkungen des MV untersuchen und in begrenztem Maße Konsequenzen aus den Erkenntnissen ziehen, die auch in die Berichte einfließen. Die Evaluation bringt demnach ausreichend Ergebnisse hervor, die als Entscheidungshilfe für Planungsmodellversuche oder bildungspolitische Entscheidungen herangezogen werden können.

Ausprägungsgrad 3:

Die Evaluationsplaner berücksichtigen die Entwicklung der Lernenden indem sie die Auswirkungen des MV untersuchen und in umfangreichem Maße Konsequenzen aus den Erkenntnissen ziehen, die auch in die Berichte einfließen. Die Evaluation bringt demnach Ergebnisse hervor, die in Art und Umfang voll als Entscheidungshilfe für Planungsmodellversuche oder bildungspolitische Entscheidungen herangezogen werden können.

K5 – Vollständige und faire Einschätzung**Definition des Standards:**

Evaluationen sollten in der Überprüfung und in der Präsentation der Stärken und Schwächen des evaluierten Modellversuchs vollständig und fair sein, so dass die Stärken weiter ausgebaut und die Problemfelder angesprochen werden können.

Operationalisierung:

Die Evaluation weist auf Stärken und Schwächen des MV hin, auch wenn diese unbeabsichtigt sind. Jedes einzelne Ergebnis ist zu begründen.

Ausprägungsgrad 0:

Nur beabsichtigte Erkenntnisse der Evaluation werden dokumentiert. Eine Begründung der Ergebnisse unterbleibt.

Ausprägungsgrad 1:

Die beabsichtigten Erkenntnisse der Evaluation werden dokumentiert und begründet, d. h. es wird deutlich, worauf die Erkenntnisse zurück zu führen sind.

Ausprägungsgrad 2:

Es werden beabsichtigte und unbeabsichtigte Erkenntnisse der Evaluation dokumentiert. Zurückgeführt auf ihre Ursachen werden jedoch nur die beabsichtigten Erkenntnisse.

Ausprägungsgrad 3:

Die Erkenntnisse (beabsichtigte und unbeabsichtigte), die die Evaluatoren gewinnen, werden im Zusammenhang dargestellt und begründet.

K6 – Offenlegung der Ergebnisse**Definition des Standards:**

Die Vertragsparteien einer Evaluation sollten sicherstellen, dass die Evaluationsergebnisse – einschließlich ihrer relevanten Beschränkungen – den durch die Evaluation betroffenen Personen ebenso wie all jenen, die einen ausgewiesenen Anspruch auf die Evaluationsergebnisse haben, zugänglich gemacht werden.

Operationalisierung:

Sämtliche Handlungen, öffentliche Erklärungen und schriftliche Berichte sind den B&B und der interessierten Öffentlichkeit zugänglich zu machen. Es werden Ansprechpartner oder eine Zentralstelle genannt, bei der alle Informationen angefordert oder eingesehen werden können.

Ausprägungsgrad 0:

Handlungen, Erklärungen und schriftliche Berichte werden den B&B sowie der interessierten Öffentlichkeit nicht zugänglich gemacht. Ein Ansprechpartner und/oder eine Zentralstelle werden nicht genannt.

Ausprägungsgrad 1:

Es werden nur sehr wenige Handlungen, Erklärungen und schriftliche Berichte (<50 %) den B&B sowie der interessierten Öffentlichkeit zugänglich gemacht. Ein Ansprechpartner und/ oder eine Zentralstelle werden nicht genannt.

Ausprägungsgrad 2:

Ein Großteil der Erklärungen, Handlungen und schriftlichen Berichte (>50 %) wird den B&B sowie der interessierten Öffentlichkeit zugänglich gemacht. Ein Ansprechpartner und/ oder eine Zentralstelle werden nicht genannt.

Ausprägungsgrad 3:

Handlungen, Erklärungen und schriftliche Berichte werden den B&B sowie der interessierten Öffentlichkeit zugänglich gemacht. Ansprechpartner und/ oder eine Zentralstelle werden genannt.

K7 – Deklaration von Interessenkonflikten**Definition des Standards:**

Interessenkonflikte sollten offen und aufrichtig behandelt werden, damit sie die Evaluationsverfahren und –ergebnisse nicht beeinträchtigen.

Operationalisierung:

Interessenkonflikte (pers. oder finanzielle Einflussfaktoren der Evaluatoren auf die Evaluation) werden im Hinblick auf ihre Ursachen dokumentiert und Maßnahmen, die vor Interessenkonflikten schützen, werden schriftlich zwischen den Betroffenen vereinbart.

Ausprägungsgrad 0:

Interessenkonflikte werden weder dokumentiert noch werden mögliche Schutzmaßnahmen vereinbart.

Ausprägungsgrad 1:

Mögliche Interessenkonflikte werden in Gesprächen ansatzweise erkannt; schriftliche Schutzvereinbarungen unterbleiben jedoch.

Ausprägungsgrad 2:

Interessenkonflikte werden in Gesprächen mit den Auftraggebern ausführlich erörtert und dokumentiert. Schutzvereinbarungen unterbleiben jedoch.

Ausprägungsgrad 3:

Vor Aufnahme der Evaluation werden Ursachen für mögliche Konflikte in Gesprächen mit dem Auftraggeber erkannt und Maßnahmen zum Schutz vor Interessenkonflikten schriftlich vereinbart.

K - Korrektheitsstandards**Definition des Standards:**

Die Korrektheitsstandards sollen sicherstellen, dass eine Evaluation rechtlich und ethisch korrekt durchgeführt wird und dem Wohlergehen der in die Evaluation einbezogenen und auch der durch die Ergebnisse betroffenen Personen gebührende Aufmerksamkeit widmet.

Operationalisierung:

Die Evaluation wird unter Beachtung rechtlicher Normen einerseits sowie gesellschaftlicher und persönlicher Werte und Normen andererseits durchgeführt. Es

ist sichergestellt, dass dem Wohlergehen der B&B ausreichend Aufmerksamkeit geschenkt wird.

Ausprägungsgrad 0:

Die Evaluation bewegt sich nicht innerhalb geltender gesetzlicher Normen und überschreitet außerdem anerkannte gesellschaftliche und persönliche Werte und Normen. Das Wohlergehen der B&B wird nicht beachtet.

Ausprägungsgrad 1:

Die Evaluation beachtet die gültige Gesetzeslage, Werte und Normen der B&B werden jedoch nicht in die Betrachtung einbezogen.

Ausprägungsgrad 2:

Die Evaluation beachtet die gültige Gesetzeslage und nimmt auch die in der Organisation anerkannten Werte zur Kenntnis. Auf die Werte des Einzelnen wird jedoch nicht eingegangen.

Ausprägungsgrad 3:

Die Evaluation beachtet die gültigen Gesetze und stellt sicher, dass dem Wohlergehen der B&B ausreichend Aufmerksamkeit geschenkt wird, indem gesellschaftliche Normen und persönliche Werte anerkannt werden.

G1 – MV-Dokumentation

Definition des Standards:

Der zu evaluierende Modellversuch sollte klar und genau beschrieben und dokumentiert werden, so dass er eindeutig identifiziert werden kann.

Operationalisierung:

Aus der Dokumentation geht hervor, wie der Modellversuch ursprünglich geplant war und wie er tatsächlich umgesetzt wurde. Aus den Beschreibungen können Inhalte und Wirkungen miteinander in Beziehung gesetzt werden.

Ausprägungsgrad 0:

Die Dokumentationen lassen nicht erkennen, wie der Modellversuch ursprünglich geplant war.

Ausprägungsgrad 1:

Die Dokumentationen lassen den originären Planungsgang nicht oder nur rudimentär erkennen; spätere Entwicklungsschritte sind, wenn überhaupt, auch nur in Grobform erkennbar.

Ausprägungsgrad 2:

Der ursprüngliche Ausgangspunkt der Planung ist überblicksartig erkennbar, die weiteren Entwicklungsschritte sind nachvollziehbar.

Ausprägungsgrad 3:

Die Dokumentationen geben den Modellversuchs-Verlaufsstand vor der Aufnahme der Tätigkeiten wieder und machen deutlich, welche Veränderungen im Laufe der Zeit vorgenommen wurden.

G2 - Kontextanalyse

Definition des Standards:

Der Kontext, in dem der Modellversuch angesiedelt ist, sollte ausreichend detailliert untersucht werden, damit mögliche Beeinflussungen des Modellversuchs identifiziert werden können.

Operationalisierung:

Der Begriff Kontext bezieht sich im Hinblick auf Planung, Durchführung und Bericht auf die Einordnung des MV in das politisch-soziale Umfeld (Kulturhoheit der Länder), den zeitlichen Rahmen, parallel laufende Aktivitäten, das Personal sowie weitere entscheidende Einflussfaktoren.

Ausprägungsgrad 0:

Das politisch-soziale Umfeld wird nicht identifiziert, andere Kontextfaktoren nicht erläutert. Eine Einordnung des MV muss somit unterbleiben. Die Nennung von Durchführungsorten, Zeitraum der Durchführung, parallel laufenden Aktivitäten, des Personals sowie weiteren entscheidenden Faktoren erfolgt nicht.

Ausprägungsgrad 1:

Das politisch-soziale Umfeld wird nicht identifiziert. Eine Einordnung des MV muss unterbleiben. Außerdem werden weitere entscheidende Faktoren nicht hervorgehoben. Durchführungsorte, Zeitraum der Durchführung, parallel laufende Aktivitäten, das Personal werden genannt.

Ausprägungsgrad 2:

Das politisch-soziale Umfeld wird identifiziert bzw. andere Kontextfaktoren werden erläutert. Eine Einordnung des MV unterbleibt jedoch. Außerdem werden weitere entscheidende Faktoren nicht hervorgehoben. Durchführungsorte, Zeitraum der Durchführung, parallel laufende Aktivitäten, das Personal werden genannt.

Ausprägungsgrad 3:

Das politisch-soziale Umfeld oder andere relevante Kontextfaktoren werden identifiziert, der MV wird hierin eingeordnet. Auch werden weitere entscheidende Faktoren hervorgehoben. Durchführungsorte, Zeitraum der Durchführung, parallel laufende Aktivitäten, das Personal werden genannt.

G3 – Beschreibung von Zielen und Vorgehen**Definition des Standards:**

Die Zwecksetzungen und das Vorgehen der Evaluation sollte ausreichend genau dokumentiert und beschrieben werden, so dass sie identifiziert und eingeschätzt werden können.

Operationalisierung:

Die Ziele der Evaluation und die beabsichtigte Nutzung der Ergebnisse werden beschrieben und vor Beginn der Evaluation den Teilnehmern mitgeteilt. Nach Abschluss des MV erfolgt eine aktuelle Beschreibung, was mit welcher Absicht gemacht wurde, damit Ergebnisse und Empfehlungen beurteilt und interpretiert werden können.

Ausprägungsgrad 0:

Es erfolgt keine Nennung der vorgesehenen Evaluationsziele und beabsichtigten Nutzungen vor, während und nach Abschluss des Modellversuchs. Eine aktuelle, begründete Beschreibung des Evaluationsvorgehens unterbleibt außerdem.

Ausprägungsgrad 1:

Die vorgesehenen Evaluationsziele und beabsichtigten Nutzungen werden genannt. Eine aktuelle und begründete Beschreibung des Evaluationsvorgehens unterbleibt jedoch.

Ausprägungsgrad 2:

Die verfolgten Evaluationsziele sowie Nutzung der Ergebnisse vor, während und nach Abschluss des MV werden genannt. Das Vorgehen bei der Evaluation wird lediglich dokumentiert.

Ausprägungsgrad 3:

Die verfolgten Evaluationsziele sowie Nutzung der Ergebnisse vor, während und nach Abschluss des MV werden genannt. Das Vorgehen bei der Evaluation wird begründend beschrieben.

G4 – Verlässliche Informationsquellen**Definition des Standards:**

Die in einer Modellversuchsevaluation genutzten Informationsquellen sollten hinreichend genau beschrieben sein, damit die Angemessenheit der Informationen eingeschätzt werden kann.

Operationalisierung:

Die genutzten Informationsquellen und die Art der Auswahl dieser Quellen wird beschrieben; die Verfahren der Informationsgewinnung werden dokumentiert und es wird deutlich, ob die Informationen einzigartig oder einseitig beeinflusst sind.

Ausprägungsgrad 0:

Die genutzten Informationsquellen werden lediglich genannt. Der Weg der Quellauswahl bleibt ebenso offen wie die Information, ob diese Quellen neutral oder einseitig beeinflusst sind.

Ausprägungsgrad 1

Die WBL nennt und beschreibt die genutzten Informationsquellen; die Art der Auswahl der Informationsquellen bleibt unklar. Die Verfahren der Datenerhebung werden nicht begründet. Außerdem bleibt offen, ob die Informationen beeinflusst sind.

Ausprägungsgrad 2:

Die Evaluatoren nennen und beschreiben die Informationsquellen, begründen deren Auswahl jedoch nicht. Die Verfahren der Datenerhebung, die auch dokumentiert sind, werden erläutert. Über eine mögliche Beeinflussung der Daten werden keine Informationen gegeben.

Ausprägungsgrad 3:

Die Evaluatoren nennen, beschreiben und begründen die Auswahl der Informationsquellen und erläutern die Verfahren der Datenerhebung, bei denen auch eine Dokumentation erfolgte. Über eine mögliche Beeinflussung der Daten wird informiert.

G5 – Valide Informationen**Definition des Standards:**

Die Verfahren zur Informationsgewinnung sollten so gewählt oder entwickelt und dann umgesetzt werden, dass die Gültigkeit der gewonnenen Interpretationen für den gegebenen Zweck sichergestellt ist.

Operationalisierung:

Es ist sicher gestellt, dass die Auswahl der Verfahren zur Informationsgewinnung begründet wird und diese derart gestaltet sind, dass die beabsichtigten Daten damit erhoben werden können.

Ausprägungsgrad 0:

Die Auswahl der Evaluationsverfahren wird nicht begründet; diese sind darüber hinaus nicht geeignet, die beabsichtigten Informationen zu gewinnen. Eine Beschreibung der Umsetzung der eingesetzten Verfahren unterbleibt ebenso, wie die Beschreibung, wie Reaktionen und Beobachtungen beurteilt oder mit Punktwerten versehen und wie Interpretationen vorgenommen wurden.

Ausprägungsgrad 1:

Die Auswahl der Evaluationsverfahren wird nicht oder nur zu einem geringen Teil begründet; außerdem sind sie nur geeignet, einen kleinen Teil der beabsichtigten Informationen zu gewinnen. Eine Beschreibung der Umsetzung der eingesetzten Verfahren unterbleibt ebenso, wie die Beschreibung, wie Reaktionen und Beobachtungen beurteilt oder mit Punktwerten versehen und wie Interpretationen vorgenommen wurden.

Ausprägungsgrad 2:

Die gewählten Verfahren zur Informationssammlung gewinnen den überwiegenden Teil der beabsichtigten Informationen, jedoch nicht alle Informationen auf die abgestellt wurde. Eine Begründung für die Auswahl der eingesetzten Evaluationsverfahren erfolgt zum überwiegenden Teil. Eine Beschreibung der Umsetzung der eingesetzten Verfahren erfolgt, jedoch fehlt die Beschreibung, wie Reaktionen und Beobachtungen beurteilt oder mit Punktwerten versehen und wie Interpretationen vorgenommen wurden.

Ausprägungsgrad 3:

Die Auswahl der eingesetzten Evaluationsverfahren wird begründet. Diese sind geeignet, die beabsichtigten Informationen zu gewinnen. Eine Beschreibung der Umsetzung der eingesetzten Verfahren erfolgt ebenso, wie die Beschreibung, wie Reaktionen und Beobachtungen beurteilt oder mit Punktwerten versehen und wie Interpretationen vorgenommen wurden.

G6 – Reliable Informationen**Definition des Standards:**

Die Verfahren zur Informationsgewinnung sollten so gewählt oder entwickelt und dann umgesetzt werden, dass die Zuverlässigkeit der gewonnenen Interpretationen für den gegebenen Zweck sichergestellt ist.

Operationalisierung:

Die Zuverlässigkeit der Evaluationsverfahren wird dadurch sichergestellt, dass bei der Auswahl der Verfahren die immanenten Zufallsfehler berücksichtigt, beschrieben und/ oder reduziert werden.

Ausprägungsgrad 0:

Hinweise auf mögliche systematische Fehler und Maßnahmen zur Reduzierung/Vermeidung dieser Fehler sind den Berichten nicht zu entnehmen.

Ausprägungsgrad 1:

Systematische Fehler werden nur ex-post benannt; Maßnahmen diese zu reduzieren oder zu beseitigen wurden nicht ergriffen.

Ausprägungsgrad 2:

Potenzielle systematische Fehler werden ex-ante erwähnt, allerdings lassen sich keine Maßnahmen zur Reduzierung bzw. Beseitigung erkennen.

Ausprägungsgrad 3:

Die Ergebnisse der Evaluation enthalten keine auf die eingesetzten Verfahren zurückzuführenden Fehler bzw. potenziell auftretende systematische Fehler werden ex-ante beschrieben und es werden Maßnahmen ergriffen, diese zu reduzieren bzw. zu beseitigen.

G8 – Analyse quantitativer Informationen**Definition des Standards:**

Quantitative Informationen einer Evaluation sollten angemessen und systematisch analysiert werden, damit die Fragestellungen der Evaluation beantwortet werden.

Operationalisierung:

Die zusammengetragenen quantitativen Daten sind nachvollziehbar geordnet, zusammengefasst, analysiert und interpretiert. Zum besseren Verständnis der Daten, der Darstellung und Interpretation der statistischen Ergebnisse sind außerdem graphische Darstellungen zu wählen. Insgesamt sind die Ergebnisse geeignet, den Einsatz der jeweiligen Evaluationsmethoden zu rechtfertigen.

Ausprägungsgrad 0:

Die zusammengetragenen Daten sind entweder nicht nachvollziehbar geordnet oder zusammengefasst worden bzw. die Aufbereitung der gesammelten statistischen Daten und Darstellung mittels Diagrammen unterbleibt ebenso wie eine umfassende Analyse und Interpretation der Daten.

Ausprägungsgrad 1:

Die zusammengetragenen Daten sind nachvollziehbar geordnet und zusammengefasst. Eine Aufbereitung der statistischen Daten mittels Diagrammen fehlt und/oder eine Analyse und/oder Interpretation unterbleibt, so dass die Ergebnisse die eingesetzten Ev.-methoden nicht/oder nur begrenzt

Ausprägungsgrad 2:

Die zusammengetragenen Daten sind nachvollziehbar geordnet, zusammengefasst und grafisch aufbereitet. Darüber hinaus werden diese analysiert und interpretiert. Die Daten lassen die eingesetzten Ev.-methoden angemessen erscheinen.

Ausprägungsgrad 3:

Die zusammengetragenen Daten sind nachvollziehbar geordnet, zusammengefasst und nachvollziehbar analysiert und interpretiert worden. Die Aufbereitung der gesammelten statistischen Daten und die Darstellung mittels Diagrammen unterstützt die vorliegenden Aussagen, aus denen auch weiterführende Schlussfolgerungen hervorgehen. Die eingesetzten Ev.-methoden werden durch die Ergebnisse legitimiert.

G9 – Analyse qualitativer Informationen

Definition des Standards:

Qualitative Informationen einer Evaluation sollten angemessen und systematisch analysiert werden, damit die Fragestellungen der Evaluation beantwortet werden.

Operationalisierung:

Die gesammelten qualitativen Informationen sind durch mehr als eine Quelle belegt. Diese Informationen sind in Kategorien unterteilt und die Schlussfolgerungen und Empfehlungen beziehen sich auf diese Kategorien. Sie sind geeignet, die Fragestellungen der Evaluation zu dokumentieren, zu erläutern und zu beantworten.

Ausprägungsgrad 0:

Die gesammelten qualitativen Informationen stammen lediglich aus einer Quelle, sind unsystematisiert und Empfehlungen beziehen sich somit auf eine unstrukturierte Datenmenge. Sollten Empfehlungen vorhanden sein, so sind diese nur begrenzt geeignet, die Fragestellungen der Evaluation zu dokumentieren, zu erläutern und zu beantworten.

Ausprägungsgrad 1:

Die gesammelten qualitativen Daten stammen aus zwei oder mehr Quellen, sind jedoch nicht in Kategorien unterteilt. Schlussfolgerungen und Empfehlungen beziehen sich somit auf eine unstrukturierte Datenmenge. Sollten Empfehlungen vorhanden sein, so sind diese nur begrenzt geeignet, die Fragestellungen der Evaluation zu dokumentieren, zu erläutern und zu beantworten.

Ausprägungsgrad 2:

Die gesammelten qualitativen Daten sind kategorisiert, jedoch beziehen sich die Empfehlungen auf die unsystematisierte Datenmenge und nicht auf die gebildeten Kategorien. Sollten Empfehlungen vorhanden sein, so sind diese nur begrenzt geeignet, die Fragestellungen der Evaluation zu dokumentieren, zu erläutern und zu beantworten.

Ausprägungsgrad 3:

Die gesammelten qualitativen Daten stammen aus mehreren Quellen (mind. zwei) und sind in Kategorien unterteilt. Die Schlussfolgerungen und Empfehlungen beziehen sich auf diese Kategorien und sind geeignet, die Fragestellungen der Evaluation zu dokumentieren, zu erläutern und zu beantworten.

G10 – Begründete Schlussfolgerungen

Definition des Standards:

Die in einer Evaluation gezogenen Folgerungen sollten ausdrücklich begründet werden, damit die B&B diese einschätzen können.

Operationalisierung:

Die Schlussfolgerungen beantworten die Fragestellungen der Adressaten, spiegeln die Verfahren und Ergebnisse der Evaluation wider; sie werden begründet und gerechtfertigt. Auch andere Erklärungen der Ergebnisse werden erarbeitet, bewertet und dargelegt.

Ausprägungsgrad 0:

Die Schlussfolgerungen gehen nicht oder nur zum Teil auf die Fragestellungen der Adressaten ein und/ oder spiegeln auch nicht die Verfahren und Ergebnisse

der Evaluation wider. Andere Erklärungen der Ergebnisse werden weder erarbeitet, noch bewertet oder dargestellt.

Ausprägungsgrad 1:

Die Schlussfolgerungen beantworten die Fragestellungen der Adressaten, spiegeln jedoch nicht die Verfahren und Ergebnisse der Evaluation wider. Andere Erklärungen der Ergebnisse werden weder erarbeitet, noch bewertet oder dargestellt.

Ausprägungsgrad 2:

Die Schlussfolgerungen gehen auf die Fragestellungen der Adressaten ein und spiegeln auch die Verfahren und Ergebnisse der Evaluation wider. Andere Erklärungen der Ergebnisse werden jedoch weder erarbeitet, noch bewertet oder dargestellt.

Ausprägungsgrad 3:

Die Schlussfolgerungen gehen auf die Fragestellungen der Adressaten ein und spiegeln auch die Verfahren und Ergebnisse der Evaluation wider. Darüber hinaus werden andere Erklärungen der Ergebnisse erarbeitet und bewertend dargestellt.

G12 – Meta-Evaluation

Definition des Standards:

Die Evaluation selbst sollte formativ und summativ in Bezug auf die vorliegenden oder andere wichtige Standards evaluiert werden, so dass die Durchführung entsprechend angeleitet werden kann und damit die Beteiligten und Betroffenen bei Abschluss deren Stärken und Schwächen gründlich überprüfen können.

Operationalisierung:

Parallel zum Evaluationsprozess des Modellversuchs findet eine Meta-Evaluation (formativ und nach Abschluss summativ) statt, die durch externe Meta-Evaluatoren durchgeführt wird.

Ausprägungsgrad 0:

Eine Meta-Evaluation wird nicht durchgeführt.

Ausprägungsgrad 1:

Es wird eine summative Meta-Evaluation durch einen Teammitarbeiter durchgeführt.

Ausprägungsgrad 2:

Sowohl eine formative als auch summative Meta-Evaluation wird durchgeführt. Der Evaluator ist Teammitarbeiter.

Ausprägungsgrad 3:

Es erfolgt eine parallele Meta-Evaluation in Form einer formativen und summativen Evaluation durch externe Evaluatoren.

G - Genauigkeitsstandards

Definition des Standards:

Die Genauigkeitsstandards sollen sicherstellen, dass eine Evaluation über die Güte und/oder die Verwendbarkeit des evaluierten Modellversuchs fachlich angemessene Informationen hervorbringt und vermittelt.

Operationalisierung:

Bei der Evaluation werden nur solche Methoden verwandt, die geeignet sind, die Güte und/oder die Verwendbarkeit des zu evaluierenden Modellversuchs zu beurteilen und auf diesem Wege für die Fragestellung relevante und ausreichende Informationen hervorbringen.

Ausprägungsgrad 0:

Die verwendeten Methoden der Evaluation sind nicht geeignet, Güte und/oder Verwendbarkeit des zu evaluierenden Modellversuchs zu beurteilen. Es werden keine für die Fragestellung relevanten Informationen hervorgebracht.

Ausprägungsgrad 1

Die verwendeten Methoden der Evaluation sind lediglich geeignet einen geringen Teil der Fragestellung zu erhellen und sind somit auch nur sehr bedingt zweckdienlich, die Güte und/oder Verwendbarkeit des zu evaluierenden Modellversuchs zu beurteilen.

Ausprägungsgrad 2:

Die verwendeten Methoden der Evaluation sind geeignet, den überwiegenden Teil der Fragestellung zu erhellen, können jedoch auch nur als bedingt zweckdienlich eingestuft werden, die Güte und/oder Verwendbarkeit des zu evaluierenden Modellversuchs zu beurteilen.

Ausprägungsgrad 3:

Die verwendeten Methoden der Evaluation beziehen sich auf die Fragestellungen des Modellversuchs, bringen die hierfür relevanten Informationen hervor und sind geeignet, die Güte und/oder Verwendbarkeit des zu evaluierenden Modellversuchs zu beurteilen.

Tabelle 4: Ausgewählte Modellversuche der Gruppe 4: Innovationen in bestehenden Fächern oder Schulformen

Förderkennziffer	Bezeichnung/Bundesland	Laufzeit
11/84-K0571	Praxisnahe Berufsausbildung im Bereich der Kommunikationsgestaltung (Baden-Württemberg)	1984-1987
52/90-K0596	Neue Entwicklungen der Informationstechnologie auf den Ausbildungsberuf Datenverarbeitungskaufmann/ Datenverarbeitungskauffrau – Auswirkungen auf Didaktik und Methodik sowie auf Lehrpläne der kaufmännischen Schulen (Baden-Württemberg)	1991-1995
3/83,83-K0640	Konzept zur Optimierung der Ausbildung an Berufsfachschulen (Bayern)	1984-1988
12/84-K0641	Erarbeitung eines pädagogisch-didaktischen Konzepts an Berufsschulen für Jugendliche, die zum Erwerb einer Berufsausbildung besonderer Lernhilfen bedürfen (Bayern)	1984-1988
41/87-K0654	Entwicklung und Erprobung besonderer Maßnahmen im Berufsschulunterricht für Schüler mit Hochschulzugangsberechtigung (Bayern)	1989-1992
4/91, 48/91 - K0665 B	Fächerübergreifender Unterricht in der Berufsschule (Bayern)	1991-1995
2/95-K0748	Initiierung und Umsetzung von Unterrichtskonzepten zum Erwerb von Schlüsselqualifikationen durch selbstbestimmtes, handlungsorientiertes Lernen unter besonderer Berücksichtigung der Lernpsychologie, der Ökologie und der Infrastruktur von berufsbildenden Oberstufenzentren (Berlin)	1995-1999
3/83;80-K0810	Entwicklung und Erprobung von Maßnahmen zur unterrichtsmethodischen Verbesserung im fachtheoretischen Unterricht der Teilzeitberufsschule (Bremen)	1984-1987
53/92, 54/92-K0863, K2863 B	Umweltbildung in Berufen der öffentlichen Verwaltung – Entwicklung und Erprobung ökologischer Lernangebote für die berufliche Aus- und Weiterbildung (Bremen)	1993 -1995
5/95-K0881	Neue Lernmethoden in der kaufmännischen Berufsausbildung (Bremen)	1995-1998
3/97, 4/97-K0889, K0889 B	Kundenorientierung in der kaufmännischen Berufsbildung (Bremen)	1997-2000
9/90-K0938	Schlüsselqualifikationen für Kaufleute im Einzelhandel (Hamburg)	1991-1993

5/97- K4019	Selbstorganisierte Lernprozesse und neue Lernwelten in der beruflichen Bildung (Hessen)	1997-2000
41/94- K2010	MV zur Entwicklung und Erprobung von komplexen Aufgaben zur Leistungsmessung und -beurteilung in der Berufsausbildung zum Bürokaufmann/ zur Bürokauffrau (Niedersachsen)	1995-1999
69/93- K2167	Differenzierung und Individualisierung in der Berufsschule für nicht neugeordnete Ausbildungsberufe (Nordrhein-Westfalen)	1994-1998
45/95, 46/95- K2181, K2181B	Integration der Europäischen Dimension in der kaufmännischen Bildung im Bildungsgang „Euro-Wirtschaftsassistent/in“ (Nordrhein-Westfalen)	1996-1999
11/88- K0240	Innere und äußere Differenzierung in Fachklassen der Berufsschule – Durchführungs- und Auswertungsphase (Rheinland-Pfalz)	1989-1992
60/91- K0251	Ganzheitliches Lernen in der Berufsschule (Rheinland-Pfalz)	1992-1995
8/91, 91/91-K 0326, K0326 B	Entwicklung und Erprobung ganzheitlicher Lernansätze in der kaufmännischen Berufsausbildung – am Beispiel des Ausbildungsberufes Versicherungskaufmann/-frau (Saarland)	1991-1994

Lebenslauf

I. Persönliche Daten

Name: Grotheer
Vorname: Michael
Geburtsdatum: 27. Januar 1968
Geburtsort: Osterholz-Scharmbeck

II. Schulischer/ Beruflicher Werdegang

1974 – 1978	Grundschule
1978 – 1980	Orientierungsstufe
1980 – 1984	Realschule
1984 – 1987	Fachgymnasium Wirtschaft
1987 – 1989	Ausbildung zum Industriekaufmann bei der Nordmende GmbH Bremen
1989 – 1990	Grundwehrdienst
1990 - 1995	Studium der Berufs- und Wirtschaftspädagogik an der Carl von Ossietzky-Universität Oldenburg in der beruflichen Fachrichtung Wirtschaftswissenschaften und dem Unterrichtsfach Deutsch
1995 – 1997	Referendariat für das Lehramt an berufsbildenden Schulen am Studienseminar Oldenburg
seit 1998	Beamter im höheren Schuldienst des Landes Niedersachsen

Osnabrück, den 06.05.2004